Development of a Technology for the Discovery of Protein Carbamates

LINTHWAITE, VICTORIA,LOUISE

Development of a Technology for the Discovery of Protein Carbamates

A thesis submitted for the degree of
Doctor of Philosophy

Victoria Linthwaite

Durham University
2017

Abstract

Carbon dioxide (CO_2) is fundamental to life with critical roles in respiration, photosynthesis, metabolism, pathogenesis, and acid-base homeostasis. It is therefore remarkable that we know so little about the direct molecular interactions of CO_2 with cellular components. CO_2 is generally unreactive but combines rapidly with neutral amines at physiological temperatures and pressures to form carbamates. Carbamylation is caused by the nucleophilic attack of an uncharged amine (lysine side chain ε-amino group or *N*-terminal α-amino group) on CO_2. The carbamate modification has been observed on proteins including RuBisCO and haemoglobin but remains largely unexplored as a protein post-translational modification.

Carbamates are labile and previous work on this PTM has involved their study under non-physiological conditions. The objective of this thesis is to investigate this understudied modification by removing its labile nature through trapping of CO_2 on its target proteins in conditions representative of a physiological environment.

This thesis presents a novel methodology to identify carbamates using a chemical trapping technique that eliminates their labile nature in combination with tryptic digest-MS analysis. The methodology functions under aqueous conditions representative of a physiological environment. Initial experiments demonstrated effective carbamate trapping at NH_2 sites within the model substrates acetyl-lysine, PHE-GLY and PHE-LYS, a tetra-peptide and haemoglobin. The results were confirmed using ESI-MS combined with ^{12}C and ^{13}C isotope incorporation. Screening of *Arabidopsis thaliana* leaf lysates identified several novel carbamylated proteins previously unknown to directly interact with CO_2. The proteomic screen was validated by the study of one new target, fructose bisphosphate aldolase 1, using recombinant protein.

This methodology provides a technology to identify sites of carbamate formation and will permit the identification of sites of CO_2 interactions within proteomes. This research has produced a method capable of removing the labile nature of carbamates and thereby completely transforming the study of carbamylation as a PTM.

Declaration

The work presented in this thesis was carried out at Durham University between October 2012 and September 2016. The work is my own original research unless otherwise indicated by statement or citation, and has not been submitted for any other qualification.

Statement of Copyright

Acknowledgments

The work in this thesis would not have been possible without a very large number of people. Firstly I would like to thank my supervisors: Dr Martin Cann for his boundless enthusiasm and Dr David Hodgson for his never ending confidence that science can be bested, even by me. I couldn't have asked for better. I would also like to thank Dr Achim Treumann and Dr Adrian Brown for their guidance and for fights against mass spec machinery.

To the members of 217 especially Louis and Peter thank you for agreeing that biology is harder than chemistry and to Hannah and Stef for your uplifting attitude and animal videos. To the members of 234 especially Phil, Chris, Alex and Alice, thank you for mostly letting me choose the radio station and the wise words that I will take with me – it's the hope that will kill you.

There really were too many friends to thank but here are some who have really earnt a mention. To begin Becky Eno and Naomi Carne for constant affirmation that we are all in this together and that really terrifying time we went sailing. To Caitlin Mooney and Arron Briddick for showing me new friends are sometimes worth the effort and all the climbing banter. To Fiona King and Rebecca Cleary first for getting me through school and since for providing a place to escape the Durham bubble when it grew too small. To Ester Harsanyi-Belteki and Lydia Harris for showing me life outside the realm of science. To *The Northerners* for letting me into your family and for KC16. To Niamh Ainsworth for your pure comedy value and not mentioning night kayaking too often. To Charlotte Blackman for your faith in whichever direction we choose to tread and the spectacular spring roll night. To Ellie Heyworth for putting up with everything (sometimes even finding it funny) and always being at the end of a phone. And to Rose Simnett, for all the days out and the nights in, I could *physically* not have got here without you.

Finally I would like to thank my family; Mum for your constant support, Dad for relinquishing the jokes about ever getting a 'real' job and Chris for enduring us all.

Thank you, it's been a blast

Contents

Abstract **2**
Declaration **3**
Statement of Copyright **3**
Acknowledgments **4**
Chapter 1: Introduction **19**
1-1 Overview 19
1-2 Physiology of CO_2 and its role *in vivo* in mammals 20
1-2.1 Carbonic anhydrase 20
1-2.2 Homeostasis of CO_2 in mammals 21
1-3 Detection pathways for CO_2 in mammals 24
1-4 Physiology of CO_2 in plants 30
1-4.1 Photosynthesis 30
1-4.2 Effect of CO_2 levels on crop yields 31
1-4.3 Homeostasis of CO_2 in plants 32
1-5 Detection pathways of CO_2 in plants 34
1-6 Carbamylation the post-translational modification 37
1-6.1 Types of Carbamylation 37
1-6.2 Carbamate formation and equilibria 38
1-7 Carbamylated proteins in mammals 41
1-7.1 Haemoglobin 41
1-7.2 Connexin 26 41
1-8 Carbamylated proteins in plants 43
1-8.1 RuBisCO 43
1-8.2 Urease 45
1-9 Carbamylated proteins in bacteria 46
1-9.1 Alanine Racemase 46
1-9.2 Transcarboxylase 46
1-9.3 β- lactamase 47
1-9.4 Phosphotriesterase 48
1-10 Practical challenge 49
1-11 Previous trapping work 51
1-11.1 TMS-DAM 51
1-11.2 Meerwein's reagents 52
1-12 Motivation for investigation 54
1-12.1 Project relevance 54

1-12.2 Aims and Hypothesis 57
Chapter 2: Materials and Methods 58
2-1 Materials and Equipment 58
2-1.1 Cell lines 58
2-1.2 pH stat 58
2-2 Experimental Biology 59
2-2.1 Trimethyloxonium tetrafluoroborate trapping experiments 59
2-2.2 Triethyloxonium tetrafluoroborate trapping experiments 60
2-3 Molecular Biology 65
2-3.1 Synthesis of acetyl-lysine carbamate 65
2-3.2 Protein digestion protocol 65
2-3.3 *Arabidopsis thaliana* plant growth 65
2-3.4 *Arabidopsis* protein extraction 66
2-3.5 Bradford assay 66
2-3.6 Bicinchoninic acid (BCA) assay 66
2-3.7 Immunocapture of RuBisCO 67
2-3.8 Ammonium sulphate precipitation 67
2-3.9 Protein solubilisation test 67
2-3.10 Acetone precipitation 67
2-3.11 Transforming cells 68
2-3.12 Protein test expression 68
2-3.13 SDS-PAGE gel 68
2-3.14 Truncation of FBA1 DNA 69
2-3.15 Polymerase chain reaction (PCR) 70
2-3.16 Agarose gel 70
2-3.17 DNA purification from agarose gel 71
2-3.18 Ligation of DNA into a vector 71
2-3.19 Mini-prep of overnight cultures 71
2-3.20 Digestion of vector and inserts 72
2-3.21 Large scale protein growth 72
2-3.22 Protein purification 72
2-3.23 One-step site directed mutagenic PCR 73
2-3.24 Refolding from inclusion bodies 73
2-3.25 FBA1 cleavage assay 74
2-4 Physical Chemistry 75
2-4.1 Concentrating a digest sample on a ziptip 75
2-4.2 Fractionating on a StageTip 75

2-4.3 Accurate mass....75
2-4.4 MALDI-MS....75
2-4.5 ESI-MS method....76
2-4.6 Data analysis....76
2-5 Synthetic chemistry....77
2-5.1 Synthesis of (But-3-enyl)dimethylsilane halogens....77
2-5.2 Synthesis of silyl propanoic acid halogens....78
2-5.3 Synthesis of amide halogens....79
2-5.4 Synthesis of amide azide....81
2-5.5 Synthesis of amide amine....82
2-5.6 Attempted synthesis of amide diazo compound....82
2-5.7 Conversion of Cl alkene to Br alkene (Finklestein reaction)....83
Chapter 3: Design and partial synthesis of a Trialkylsilyl-diazomethane Derivative for Carbamate Trapping....84
3-1 Overview....84
3-2 Background to the trapping hypothesis....85
3-2.1 Diazomethane....85
3-2.2 Trimethylsilyl-diazomethane....86
3-3 Design of a water-soluble TMS-DAM derivative....89
3-4 Retrosynthetic strategy for the chemical synthesis of designed carbamate trapping agent 1....92
3-5 Forward Synthesis of Carbamate trapping agent 1....94
3-5.1 Synthesis of Alkenes....94
3-5.2 Oxidation of alkenes 4 to acids 5....96
3-5.3 Formation of amides 9 from acids 5....99
3-5.4 Diazo group transfer to halo amides 9....103
3-5.5 Interconversion of halo substituents of alkenes 4 via the Finkelstein reaction 104
3-5.6 Halide to azide conversion....106
3-5.7 Conversion of azide 10 to diazo 1....109
3-5.8 Conversion of azide 10 to amine 14 for conversion to diazo 1....112
3-6 Conclusion....113
3-7 Future work....114
Chapter 4: Development of a carbamate trapping method....115
4-1 Overview....115
4-2 Methods to modify carboxyl groups and previous work with Meerwein's reagent....116
4-2.1 Trimethyloxonium tetrafluoroborate (TMO)....117
4-2.2 Triethyloxonium tetrafluoroborate (TEO)....118

4-3 Carbamate trapping using TMO....121
4-3.1 Phenylacetic acid....121
4-3.2 Acetyl-lysine....123
4-3.3 Amount of possible carbamate formation....128
4-4 Trapping carbamates using TEO....129
4-4.1 Acetyl-lysine trapping....129
4-4.2 Synthesis of synthetic carbamylated acetyl-lysine....132
4-4.3 ^{13}C Confirmation of carbamylated acetyl-lysine....134
4-4.4 Phenylalanine trapping....135
4-5 Development of the trapping method on dipeptide and tetra peptide systems....139
4-5.1 Glycine- Phenylalanine (GLY- PHE)....139
4-5.2 Lysine- Phenylalanine (LYS- PHE)....140
4-5.3 Phenylalanine-leucine-lysine-glutamine (FLKQ) Tetrapeptide....142
4-6 Haemoglobin trapping....144
4-6.1 Proteomics....145
4-6.2 MALDI....146
4-6.3 ESI....149
4-7 Conclusion....153
4-8 Future work....154
Chapter 5: A proteomic screen for new protein carbamates....155
5-1 Overview....155
5-2 Previous work on complex protein systems....156
5-2.1 Previous proteome investigations....156
5-2.2 Previous work investigating PTMs....158
5-3 Development of a method to identify unique proteins from a carbamate trapping reaction....160
5-3.1 Analysis of amount of protein extracted from *Arabidopsis* leaves....160
5-3.2 Fractionation of the sample....162
5-3.3 Solubilisation....165
5-3.4 Fractionation of peptides....168
5-4 Analysing protein carbamates....171
5-4.1 Difficulty in analysing data....171
5-4.2 Software used for analysis....172
5-4.3 Carbamate hits found....175
5-4.3.1 Tandem....175
5-4.3.2 MaxQuant....182
5-4.4 False positive examples....185

5-4.5 Summary of total hits 188
5-5 Conclusion 190
5-6 Future work 191
Chapter 6: Expression and validation of FBA1 193
6-1 Overview 193
6-2 Factors in recombinant protein expression 194
6-2.1 Choice of organism 194
6-2.2 Choice of cloning vector 194
6-2.3 Purification tags 196
6-3 Expression of FBA1 197
6-3.1 Test expression 197
6-3.2 Large scale protein expression 200
6-3.3 Inclusion bodies 202
6-3.4 Improved soluble protein conditions 203
6-3.5 Soluble FBA1-WT-Trunc purification 204
6-3.6 Construction of mutant FBA1 protein 206
6-4 Purified FBA1 carbamate trapping 208
6-5 Fructose Bisphosphate aldolase 1 assay 209
6-5.1 Aldolase reaction mechanism 209
6-5.2 Synthesised FBA1-WT and FBA1-K293A activity assay 210
6-5.3 CO_2 dependence on activity 211
6-5.4 Protein structural information 213
6-6 Conclusion 215
6-7 Future work 216
Chapter 7: Synopsis 217
7-1 Introduction 217
7-2 Partial synthesis of a carbamate trapping reagent 218
7-3 Development of the trapping methodology using Meerwein's reagents 219
7-4 Newly discovered carbamate sites 220
7-5 Validation of a previously unknown protein carbamylation site 222
7-6 Conclusions 223
7-7 Future work 224

Abbreviations

1,3-BPG	1,3-bisphosphoglyceric acid
2-DE	2D gel electrophoresis
3-PG	3-phosphoglycerate
ABA	Abscisic acid
ABC	Ammonium bicarbonate buffer
AQP	Aqua-porin
Arabidopsis	*Arabidopsis thaliana*
ARDS	Acute respiratory distress syndrome
AS	Active site
Ar	Argon
BCA	Bicinchoninic acid assay
Br	Bromine
BSA	Bovine serum albumin
CA	Carbonic anhydrase
Ca^{2+}	Calcium ions
cAMP	Cyclic adenosine monophosphate
CBB	Calvin Benson Bassham
cGMP	Cyclic guanidine monophosphate
Cl	Chlorine
CO_2	Carbon dioxide
Cu^{2+}	Copper ions
Cx26	Connexin 26
Da	Daltons
DAG	Diacylglycerol
DAM	Diazomethane
DCC	Dicyclohexylcarbodiimide
DCM	Dichoromethane

dCO_2	Dissolved carbon dioxide
DCU	Dicyclohexyurea
DHAP	Dihydroxyacetone phosphate
DPPA	Diphenyl phosphorazidate
DTT	Dithiothreitol
E. coli	*Escherichia coli*
EDTA	Ethylenediaminetetraacetic acid
ER	Endoplasmic reticulum
ESI	Electrospray ionisation
F-1,6-BP	Fructose-1,6-bisphosphate
FACE	Free-air concentration enrichment
FBA1	Fructose bisphosphate aldolase 1
FLKQ	Phenylalanine leucine lysine glutamine
G3P	Glyceraldehyde 3-phosphate
GPDH	α-glycerophosphate dehydrogenase
GOI	Gene of interest
GPCRs	G-protein coupled receptors
GTP	Guanosine triphosphate
Hb	Haemoglobin
HCl	Hydrochloric acid
HCO^{3-}	Bicarbonate ion
HMGB1	High mobility group box 1
IKKα	Inhibitor of κB
IL-8	Interleukin 8
IP_3	Inositol 1,4,5-triposphate
IP_3R	IP_3 receptor
IPTG	Isopropyl β-D-1-thiogalactopyranoside
IRBIT	IP_3R binding protein
K^+	Potassium ion

KCl	Potassium chloride
LB	Luria broth
LC-ESI-MS	Liquid chromatography-electrospray-mass spectrometry
LC-MS	Liquid chromatography mass spectrometry
LiBr	Lithium bromide
LTP	Lipid transfer protein
MALDI	Matrix assisted laser absorption ionisation
MeCN	Acetonitrile
$MgCl_2$	Magnesium chloride
$MgSO_4$	Magnesium sulphate
MS	Mass spectrometry
MS-MS	Tandem mass spectrometry
MT	Mutant
MW	Molecular weight
NaBr	Sodium bromide
NaCl	Sodium chloride
NAD^+	Nicotinamide adenine dinucleotide
NADH	Reduced nicotinamide adenine dinucleotide
NADPH	Reduced nicotinamide adenine dinucleotide phosphate
$NaHCO_3$	Sodium bicarbonate
NaOH	Sodium hydroxide
NBC1	Sodium/bicarbonate cotransporter
NMR	Nuclear magnetic resonance
NO	Nitric oxide
O_2	Oxygen
OH^-	Hydroxide ion
ori	Origin of replication
OSN	Olfactory sensory neuron
OSNs	Olfactory sensory neurons

PCR	Polymerase chain reaction
PHE-GLY	Phenylalanine-glycine
PHE-LYS	Phenylalanine-lysine
pH*i*	Intracellular pH
pI	Isoelectric point
PIP2	Phosphatidylinositol 4,5-bisphosphate
PKA	Protein kinase A
PKG	Protein kinase G
PLC	Phospholipid C
pNBC1	Pancreas NBC1
POI	Protein of interest
PTMs	Post-translational modifications
PVPP	Poly(vinylpolypyrolidone)
RBC	Red blood cell
RNAP	T7 RNA polymerase
ROS	Reactive oxygen species
RuBisCO	Ribulose-1,5-bisphosphate carboxylase/oxygenase
RuBP	Ribulose 1,5-bisphosphate
sAC	Soluble adenylyl cyclase
SAP	Shrimp alkaline phosphatase
SCX	Cation exchange
SDS	Sodium dodecyl sulfate
SDS-PAGE	Sodium dodecyl sulfate-polyacrylamide gel electrophoresis
sGC	Soluble guanadyl cyclase
SLAC1	Slow anion channel-associated 1
SN	Supernatant
Tandem	GPM X!Tandem
TBE	Tris/Borate/EDTA buffer
TEO	Triethyloxonium tetrafluoroborate

TFA	Trifluoroacetic acid
TIC	Total ion chromatogram
tmAC	Transmembrane adenylyl cyclase
TMO	Trimethyloxonium tetrafluoroborate
TMS-DAM	Trimethylsilyl diazomethane
TPI	Triosephosphate isomerase
UV	Ultraviolet
WT	Wild-type
α-CHCA	Alpha-cyano-4-hydroxycinnamic acid

List of Schemes

Scheme 1-1 Nucleophilic attack of a neutral amine on CO_2 to form a carbamate 19

Scheme 1-2 Equilibria between the hydration of CO_2 to bicarbonate 20

Scheme 1-3 Process of hydration of CO_2 by CA active site 21

Scheme 1-4 The reversible nucleophilic attack of a neutral amine on CO_2 38

Scheme 1-5 The RuBisCO active site 44

Scheme 1-6 Labile reaction mechanism of carbamate formation 49

Scheme 1-7 Removing the labile nature of a carbamate 50

Scheme 1-8 Esterification of a generic carbamate by TMS-DAM 51

Scheme 1-9 Ethylation of a carbamate by TEO 53

Scheme 1-10 Hydrolysis of Meerwein's reagent TEO 53

Scheme 3-1 Trapping of a carbamate using an electrophilic reagent 84

Scheme 3-2 Resonance structures of diazomethane 85

Scheme 3-3 Mechanism of esterification of a carboxylic acid by diazomethane 86

Scheme 3-4 In situ formation of DAM from TMS-DAM 87

Scheme 3-5 Hypothesised mechanism of trapping of a carbamate 90

Scheme 3-6 Retrosynthetic scheme for synthesis of diazo trapping molecule 1 92

Scheme 3-7 Synthesis of alkenes 4 94

Scheme 3-8 Synthesis of carboxylic acids 5 by oxidation of haloalkenes 4 97

Scheme 3-9 Reaction mechanism of the formation of amides 9 100

Scheme 3-10 Synthesis of amide 9 101

Scheme 3-11 Attempted synthesis of diazo molecule 1 103

Scheme 3-12 Synthesis of bromo alkene 4 (X=Br 105

Scheme 3-13 Literature method of converting a chlorine to an azide group 107

Scheme 3-14 Synthesis of silyl azide 10 107

Scheme 3-15 Synthesis of silyl amine 13 112

Scheme 4-1 Modification of a carboxyl group by different alkylation methods 116

Scheme 4-2 Hypothesised reaction mechanism of methyl transfer from TMO 117

Scheme 4-3 Hydrolysis of TEO reagent producing H^+ ions 119

Scheme 4-4 Transfer of an ethyl group from TEO to carboxylic acid side chain 119

Scheme 4-5 Transfer of a methyl group from TMO to phenylacetic acid 121

Scheme 4-6 Dissociation of sodium bicarbonate to carbon dioxide 123

Scheme 4-7 The process of carbamate formation on lysine 124

Scheme 4-8 Equilibria present with acetyl-lysine in solution 127
Scheme 4-9 Transfer of an ethyl group from TEO to a carbamate 129
Scheme 4-10 The process of carbon dioxide binding to lysine 130
Scheme 4-11 Synthesis of chemically modified acetyl-lysine derivative 133
Scheme 4-12 Formation and trapping of a carbamate on phenylalanine 136
Scheme 4-13 PHE-GLY dipeptide reaction with CO_2 and TEO 139
Scheme 4-14 PHE-LYS dipeptide and its possible carbamate formation sites 141
Scheme 4-15 FLKQ tetrapeptide reaction with CO_2 and TEO 142
Scheme 4-16 Trapping of a carbamate formed Hb 145
Scheme 5-1 Alkyne-azide click chemistry to modify the TEO trapping reagent 191
Scheme 7-1 Nucleophilic attack of an uncharged amine on carbon dioxide 217
Scheme 7-2 Process of ethyl transfer from TEO to a carbamate 219

List of Figures

Figure 1-1 Simplified process of CO_2 entrance within a RBC 23
Figure 1-2 Interaction between the cAMP and cGMP pathways 25
Figure 1-3 Activation of the IP3 pathway by PIP2 hydrolysis 26
Figure 1-4 Interaction of IRBIT with IP3R to regulate the Na^+/HCO_3^- cotransporter 27
Figure 1-5 The Calvin Cycle 30
Figure 1-6 Pathways involved in the process of stomata closing and opening 35
Figure 1-7 Difference between the isocyanate carbamate and CO_2 carbamate. 37
Figure 1-8 Equilibria necessary for carbamate formation 39
Figure 1-9 Meerwein's reagents TMO and TEO 52
Figure 3-1 Structure of TMS-DAM 86
Figure 3-2 Silyl diazo molecule showing introduction of a water-soluble group 89
Figure 3-3 Design of water soluble carbamate trapping agent 1 91
Figure 3-4 Comparison of 4-bromo-1-butene and halo alkene 4(X=Br) 96
Figure 3-5 1H NMR comparison of the bromo alkene 4 and Br acid 4 98
Figure 3-6 13C NMR spectra of bromoalkene 4(X=Br) and acid molecule 5(X=Br) 99
Figure 3-7 1H NMR spectrum for bromo amide 9 102
Figure 3-8 The trans and cis forms of the amide bond formation for amide 9 103
Figure 3-9 1H NMR showing movement of signals from 4(X=Cl) to 4(X=Br) 105

Figure 3-10 1H NMR spectra for azide amide molecule 10 108
Figure 3-11 Phosphine 11 109
Figure 3-12 Phosphine molecule binding to the azide functional group 110
Figure 4-1 Meerwein's reagents TMO and TEO 115
Figure 4-2 1H NMR synthesised phenylacetate and purchased phenylacetate 122
Figure 4-3 13C NMR spectrum formation of a carbamate on acetyl-lysine 124
Figure 4-4 ESI-MS acetyl-lysine with a methylated carboxyl group 125
Figure 4-5 Acetyl-lysine with additional methyl group transfer 126
Figure 4-6 Henderson-Hasselbalch equation 128
Figure 4-7 Graph produced by pH stat 131
Figure 4-8 TIC and m/z trace from trapping of acetyl-lysine with TEO 132
Figure 4-9 1H NMR chemically carboxyl ethylated acetyl-lysine 133
Figure 4-10 MS trace showing 12C vs 13C 135
Figure 4-11 TIC and UV traces of trapped phenylalanine reaction at pH 7.4 137
Figure 4-12 Comparison of UV spectra phenylalanine at pH 7.4 vs pH 8.5 (B) 140
Figure 4-13 ESI-MS showing the trapped carbamate product of PHE-GLY 141
Figure 4-14 ESI-MS showing the trapped carbamate product of PHE-LYS 143
Figure 4-15 ESI-MS showing the trapped carbamate product of FLKQ 144
Figure 4-16 13C NMR spectrum showing a carbamate peak 147
Figure 4-17 Hb MALDI spectrum 148
Figure 4-18 MSMS fragmentation peaks of the N-terminal peptide peak 1080.6 150
Figure 4-19 Tandem image of N-terminal peptide with a trapped carbamate 151
Figure 4-20 Tandem image of N-terminal peptide with a 13C trapped carbamate 152
Figure 5-1 Steps involved in bottom-up proteomics for peptide identification 156
Figure 5-2 Standard curve from the Bradford assay 160
Figure 5-3 Standard curve from the BCA assay. 161
Figure 5-4 SDS-PAGE of a RuBisCO depleted sample 162
Figure 5-5 SDS-PAGE of four salt precipitation samples 163
Figure 5-6 Number of unique proteins found in each salt precipitation sample 164
Figure 5-7 Ethylation of acidic amino acids (D and E) by TEO 165
Figure 5-8 Amount of protein present before trapping and after trapping 166
Figure 5-9 Total amount of protein from four solubilisation methods 167
Figure 5-10 Comparison of pellet solubilisation with increased ABC buffer 168
Figure 5-11 Comparing number of unique proteins using sample fractionation 169
Figure 5-12 Comparing number of unique proteins between MS machines 170

Figure 5-13 Exact mass of carbamate addition on a lysine side-chain 175
Figure 5-14 12C vs 13C carbamate trapped on K44 of the lipid transfer protein 177
Figure 5-15 Software displays of the K183 carbamate hit found on RuBisCO 179
Figure 5-16 Software displays of the K262 carbamate hit found on Peroxidase 180
Figure 5-17 Tandem result showing a trapped carbamate at K293 on FBA1 181
Figure 5-18 Peptide containing the K109 carbamate on PSII subunit Q 183
Figure 5-19 Peptide containing the K696 carbamate on Tetratricopeptide repeat 184
Figure 5-20 Example of a false positive carbamate site. 185
Figure 5-21 Bad fragmentation pattern not around the carbamate modification 187
Figure 6-1 pET-14b expression vector 195
Figure 6-2 Purification of POI containing a His tag 196
Figure 6-3 FBA1 protein sequence 197
Figure 6-4 SDS-PAGE of test expression of FBA1 198
Figure 6-5 SDS-PAGE of Rosetta strain test 199
Figure 6-6 SDS-PAGE of BL21 strain test 201
Figure 6-7 SDS-PAGE of refolding of FBA1-WT from inclusion bodies 202
Figure 6-8 SDS-PAGE of test conditions to improve of soluble FBA1-WT 203
Figure 6-9 SDS-PAGE of purification of FBA1-WT-Trunc using a Ni^{2+} column 205
Figure 6-10 SDS-PAGE of improved purification of FBA1-WT-Trunc 206
Figure 6-12 Aldolase reaction mechanism 209
Figure 6-13 Coupling aldolase reaction mechanism to produce enzyme assay 210
Figure 6-14 Graph of enzyme assay results 211
Figure 6-15 Model structure of FBA1 213
Figure 7-1 Diazo trapping molecule **1** 218
Figure 7-2 Synthetic final product from Chapter 3 218

List of Tables

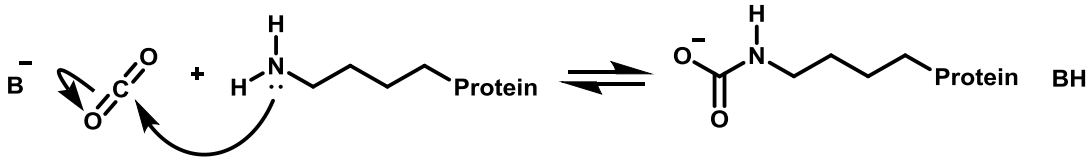
Table 2-1 Components and conditions for the amplification of purified DNA 70
Table 3-1 Optimisation experiments towards the preparation of haloderivatives 4 95
Table 3-2 Details of method development for the oxidation of alkenes 4 to acids 5 97
Table 3-3 Details of development process of synthesis of amide 9 101

Table 3-4 Details of development of addition of amine 6 to acid 5 102
Table 3-5 Details of attempts towards diazo group transfer to amide 9 104
Table 3-6 Details of the optimism of the Finkelstein reaction 106
Table 3-7 Details of development process 108
Table 3-8 Summary of workup conditions for the breakdown of acyl triazene 111
Table 4-1 Showing the corresponding modifications for the MS in Figure 4-17 146
Table 4-2 Comparing the fragments from 1080.6 N-terminal peptide sequence 148
Table 5-1 Variable modifications that can occur during the reaction with TEO 172
Table 5-2 Combined carbamate hits found using both analysis software 188

Chapter 1: Introduction

1-1 Overview

The research presented within this thesis is an investigation into the modification of proteins by CO_2. This reaction occurs by the nucleophilic attack of a neutral amine on CO_2 to form a post-translational modification known as a carbamate (Scheme 1-1).



Scheme 1-1 The reversible nucleophilic attack of a neutral amine on CO_2 to form a carbamate.

Carbamates have previously been understudied due to their labile nature and subsequently there is little known about their interaction within most biological systems. The importance of investigating carbamates is to provide further insight into the molecular mechanisms of CO_2 within a cellular system and the possible effects that either an elevated or reduced level could cause.

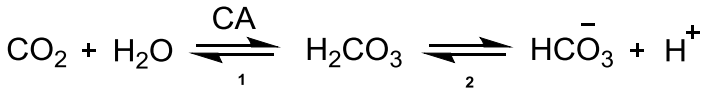
This work aimed to create a tool which could be used in a model physiological environment to reduce the labile nature of a carbamate and allow for downstream analysis. This tool could then be used to screen biological systems for undiscovered carbamates.

The aim of this chapter is to provide a general overview of CO_2, the effects it has in systems, the carbamate modification itself and the reason that studying CO_2 is so important.

1-2 Physiology of CO_2 and its role *in vivo* in mammals

1-2.1 Carbonic anhydrase

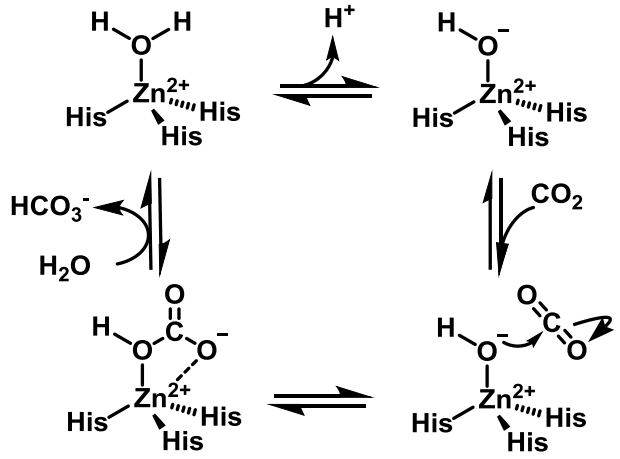
CO_2 has fundamental roles in all biological systems via its effect on photosynthesis (1), respiration (2) and acid-base homeostasis (3). CO_2 in solution reacts with water molecules to produce carbonic acid which rapidly breaks down to bicarbonate ions and protons (Scheme 1-2). This system connects CO_2 to the pH of a solution (4).



Scheme 1-2 Equlibria between the hydration of CO_2 to carbonic acid by CA (1) and the breakdown of carbonic acid to bicarbonate (2).

The limiting equilibrium in this process is the hydration of CO_2 (**1**). The actual process of water and CO_2 combining is too slow at physiological pH (5,6) and is therefore catalysed by carbonic anhydrase (CA) (7). Carbonic anhydrase enzymes are very efficient enzymes with a turnover rate of up to 1 million CO_2 molecules per second (8). Their role is important in all systems that involve CO_2 such as during photosynthesis and respiration. This process is so important that CA enzymes have been found in Archaea, prokaryotes and eukaryotes (9). The equilibria in Scheme 1-2 mean that changes in the concentration of any of these components reflect the concentration of all three (10). This relationship relates the pH of the system and CO_2 homeostasis (9).

CA is a zinc-containing enzyme with three evolutionary unrelated families known as α, β and γ. The mechanism of action involves an OH^- ion bound to a coordinated zinc molecule attacking CO_2 to form a metal-bound HCO_3^- ion which is displaced by a water molecule (7) (Scheme 1-3). Animals only contain the α-type of CA but higher plants and cyanobacteria contain members of all three families (11).



Scheme 1-3 Hydration of CO_2 by CA active site.

1-2.2 Homeostasis of CO_2 in mammals

The physiology for the regulation of CO_2 within the aveolar system was thoroughly covered by Haldane and Priestley 1905 where they investigate the effects of oxygen deficiency or excess carbon dioxide on lung ventilation (12). This paper was also the first to suggest the modification of a protein (Hb) by carbon dioxide causing an affect on the function of the protein. This is now known to be by carbamate formation. Further from this work it is now known that homeostasis of internal CO_2 levels is fundamental for all kingdoms of life (13). The control of both blood gases CO_2 and O_2 are important within systems, with P_{CO2} regulation hypothesised to be more important than P_{O2} due to the tighter control over levels of CO_2 (14). Without homeostasis cellular systems could undergo respiratory poisoning and pH levels could move outside of the range permitted for enzyme reactions (13), a movement from this range of just 0.1 μM can be fatal (15). As shown in Scheme 1-2 CO_2 equilibria is directly linked to system pH by the formation of H^+ ions. pH levels are important for many chemical and enzymatic processes within a cell such as cationic ion channel opening (16), meaning regulation of the intracellular CO_2 equilibrium is essential. The bicarbonate system is viewed as the most important buffer system in

the body (17) and its levels are closely maintained by reabsorption and excretion (18).

Cells transport H^+ and HCO_3^- across cell membranes via cationic and anionic channels to maintain these levels (10). Acid-base transport across the plasma membrane is crucial for intracellular pH (pH*i*) control. Disturbed levels can cause ion channel malfunction and affect enzyme activity (19). There are several G protein-coupled receptors (GPCRs) known to be activated by a reduction in pH (below 6.8) caused by H^+ levels (10). Conversion between these species is also important for transport, cells convert H_2CO_3 to CO_2 which can then diffuse into the renal cell as HCO_3^- cannot pass through the cell membrane (5). The detection of CO_2 in a system can also be through monitoring of the production of HCO_3^- and H^+ which then triggers signal cascades which are monitored (20).

Despite its importance, very little is actually understood of the molecular interactions of CO_2 within the cell. It was originally assumed that all gases were transported by passage through a biological membrane by diffusion. Gutknecht *et. al.* 1977 found that lipid permeability to CO_2 was greater than that of H_2CO_3 due to polarity (3). However it was later shown that the partition coefficient of CO_2 was reduced slightly in the presence of cholesterol (21). It is now agreed that some membranes are more permeable than others and in fact, some membranes such as the apical membranes of gastric gland cells and colonic crypt cells, are not permeable to CO_2 at all (22,23). In some cells, such as red blood cells, it is now known that this process is helped by transporter proteins such as Rhesus and aquaporin proteins (AQP) (24). A study undertaken using aquaporin-1 (AQP1) knockout mutants was able to show that the level of CO_2 inside a red blood cell (RBC) compared to outside was significantly reduced upon lack of AQP1 channel (22). Within the blood system CO_2 is transported either dissolved in solution, as carbonic acid or bound to haemoglobin as a carbamate (25,26). The transport of CO_2 by the blood therefore involves the hydration of CO_2 to carbonic acid and then to a bicarbonate ion. After the hydration reaction of the CO_2 the carbonic anhydrase dissociates (27), allowing the movement of bicarbonate across the membrane in exchange for a Cl^- ion (Figure 1-1).

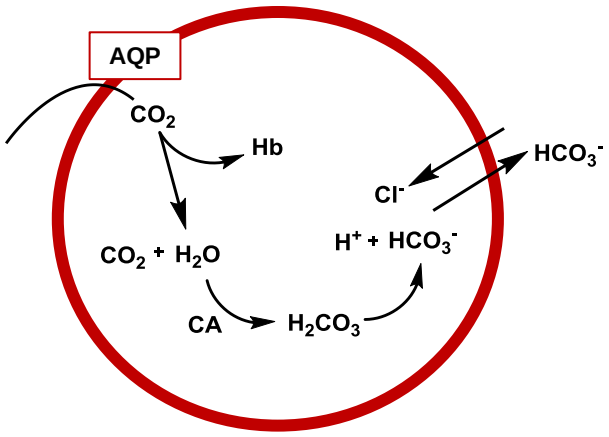


Figure 1-1 Simplified process of CO_2 entrance, equilibria within and then exit from a RBC.

The blood plasma of the average adult contains 50-65 % of its CO_2 gas as bicarbonate (28). It is clear from the described equilibrium that an increase in the amount of CO_2 present will correlate with an increase in the amount of bicarbonate (28).

1-3 Detection pathways for CO_2 in mammals

CO_2 and HCO_3^- influence a number of cell signalling processes. Identification of CO_2 responsive signalling pathways is key to understanding how organisms cope with changing CO_2 levels.

Cyclic adenosine 3',5',monophosphate (cAMP) is a secondary cell messenger present in many signalling cascades. cAMP is regulated by two adenylyl cyclase forms, one is a transmembrane enzyme (tmAC) that is activated by a G protein and one is soluble adenylyl cyclase (sAC). sAC is directly regulated by the presence of HCO_3^- (29). This regulation shows a direct link from CO_2 to cellular signalling pathways (30).

The cAMP produced by adenylyl cyclase is responsible for activating protein kinase A (PKA) which phosphorylates downstream proteins. One consequence of this pathway is the release of calcium ions (Ca^{2+}) from the ER by phosphorylation of calcium channels. These calcium ions are messengers in numerous pathways. One of which is the production of cGMP via calcineurin and nitric oxide (31). cGMP activates protein kinase G (PKG) (Figure 1-2).

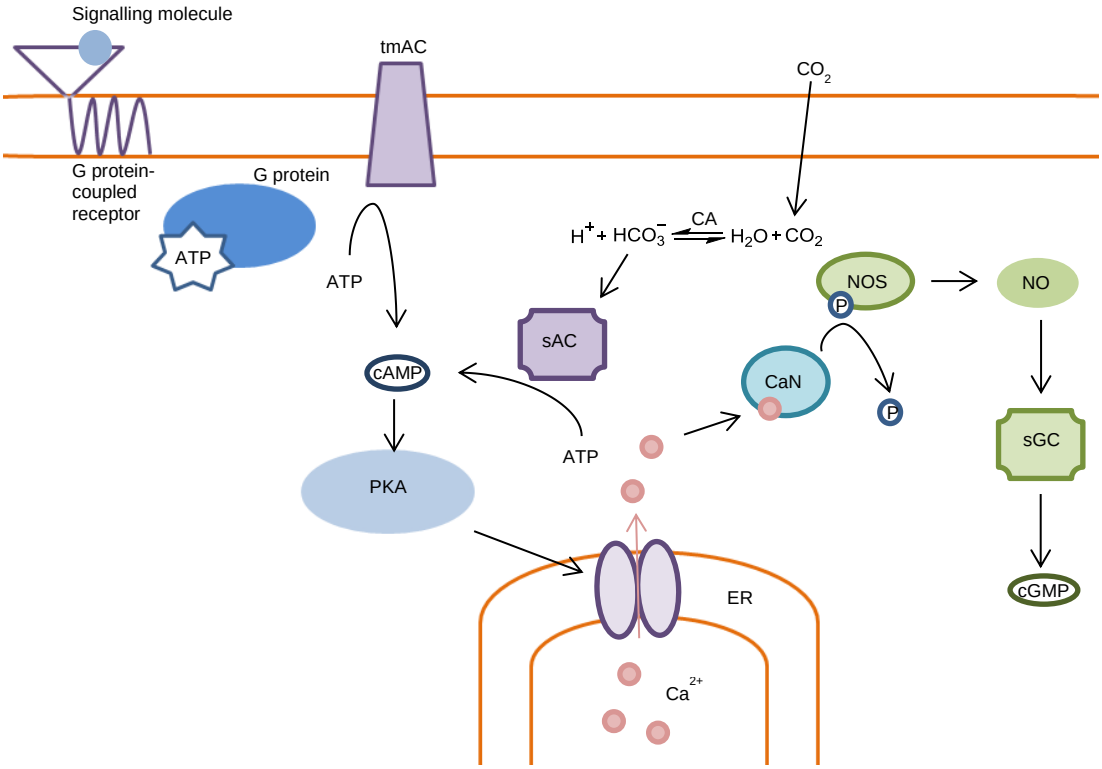


Figure 1-2 Interaction between the cAMP and cGMP pathways showing intracellular calcium signalling. Both cAMP and cGMP are produced by soluble cyclase enzymes. Protein kinase A (PKA), soluble adenylyl cyclase (sAC), calcineurin (CaN), nitric oxide species (NOS), soluble guanadylyl cylase (sGC).

Townsend *et. al.* 2009 demonstrated that some subclass members of AC Class III respond to CO_2 instead of HCO_3^-. The phylogenetically related Class IIIa Rv1625c AC of *M. tuberculosis* H37Rv was also shown to be directly stimulated by CO_2 at physiological concentrations. This led to the conclusion that the mammalian cAMP signalling pathway is able to discriminate between CO_2 and HCO_3^- *in vivo*.

IP_3 receptors (IP_3R) are gated Ca^{2+} channels that allow the release of Ca^{2+} into the intracellular environment (32). These receptors are activated by the hydrolysis product of phosphatidylinositol 4,5-bisphosphate (PIP_2). PIP_2 is cleaved to form diacylglycerol (DAG) and IP_3, this IP_3 goes on to activate the IP_3R which leads to the release of Ca^{2+} from the ER (33) (Figure 1-3). This calcium release regulates thousands of downstream targets.

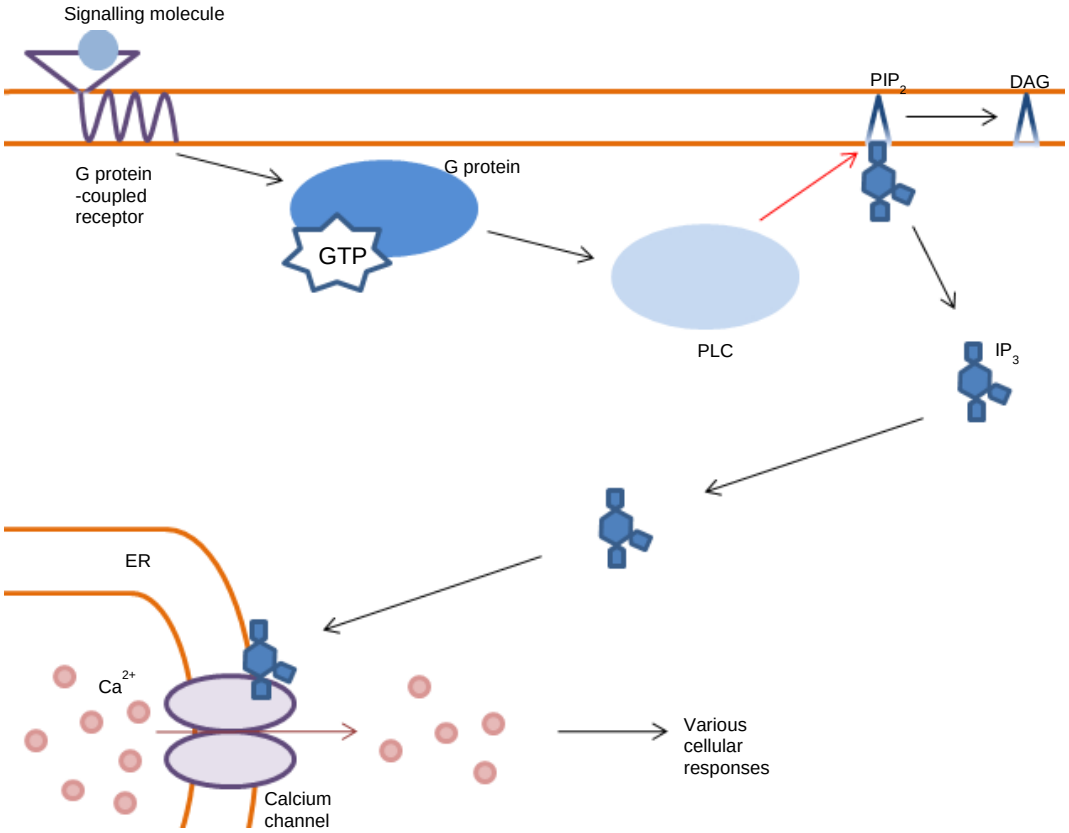


Figure 1-3 Activation of the IP_3 pathway by PIP_2 hydrolysis causing Ca^{2+} release from the ER.

The activation of IP_3R also causes the release of the inositol 1,4,5-triphosphate receptor-binding protein (IRBIT), which is an alternative binding ligand for IP_3R able to compete with IP_3 to suppress IP_3R and regulate Ca^{2+} release (34).

NBC1 is a cotransporter that facilitates the movement of Na^+ and HCO_3^- ions across plasma membranes, with 2 or 3 HCO_3^- for each Na^+ (34). This links NBC1 to the net movement of negative charge across the membrane. IRBIT has been found to specifically bind to pNBC1 (pancreas NBC1) cotransporters and increase their activity (Figure 1-4) (35). This links IRBIT to pH regulation and shows a method of regulating variants of NBC1 for different physiological roles (33).

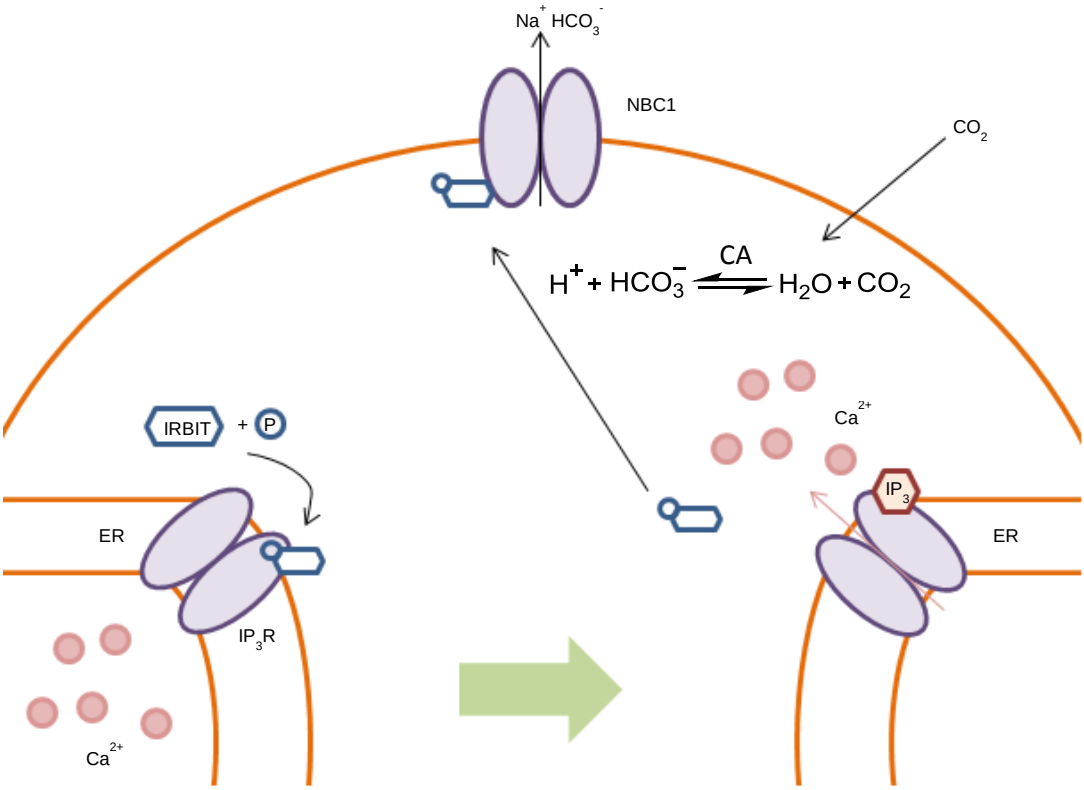


Figure 1-4 Interaction of IRBIT with IP_3R to regulate the Na^+/HCO_3^- cotransporter NBC1 and cause calcium release.

IP_3R is also regulated by phosphorylation by PKA linking the removal of intracellular HCO_3^- to the cAMP pathway (36). This introduces an insight into the crosstalk between the cAMP and Ca^{2+} release pathways (37). These pathways also link in the regulation of IP_3 production. The phosphorylation of PLC by PKA has been shown to prevent IP_3 production by PIP_2 hydrolysis (37). Cytosolic IP_3 levels are also regulated by their conversion into IP_2 and IP_4 by IP_3 5-phosphatase and IP_3 3-kinase respectively.

Reactive oxygen species (ROS) have been shown to be important chemicals in signalling processes involving cell growth and differentiation (38). Nitric oxide (NO) has been identified as a mediator in the immune response (39) via the production of sGC which catalyses cGMP production (Figure 1-2). ROS can reduce damage by ischemic injury to liver cells by activating the release of HMGB1 via calcium dependent kinases (40). However hypoxia caused ROS formation can cause cellular damage. ROS are produced by NADPH oxidase activity which has been shown to be inhibited by CO_2 (41).

It has recently been shown that similar to ROS and O_2, CO_2 is also sensed by cells causing physiological responses via NF-$_K$β transcription factors (42). NF-$_K$β is a transcription factor that regulates genes responsible for the immune response. In lung cells it was discovered that hypoxia-mediated IL-8 induction was as a result of NF-$_K$β activation (43). The mechanism of this reaction is caused by the methylation of phosphatase PP2A which controls the translocation of NF-$_K$β to the nucleus. Guais *et. al.* 2011 found that this reaction was not induced under a drop in pH without the presence of CO_2 (44) implying the interaction of CO_2 in this pathway. Mammalian cells sense changes in CO_2 levels leading to altered gene expression via the NF-$_K$β pathway (45). IKKα is a regulatory component of NF-$_K$β which has been found to translocate to the nucleus in response to elevated CO_2. The results of this imply the existence of a molecular CO_2 sensor in mammalian cells.

Chemosensing is a complex mechanism of signal transductions that relate to a cellular response at a biological level (46). Cells that are sensitive to changing levels of CO_2/H^+ are chemosensitive cells (47). These cells are necessary to maintain homeostatic control over blood gas concentrations (48). The specific mechanisms of this process are unclear but the pathways to detect CO_2 are known to be linked by sAC regulation through bicarbonate sensing (49). sAC then regulates downstream pathways via the secondary messenger cAMP. There are several types of chemoreceptor, they may sense CO_2, HCO_3^- or pH (10) in order to regulate breathing and tidal volumes. The mechanism of conversion of bicarbonate into CO_2 is important for several reasons. For example high dissolved CO_2 (dCO_2) concentrations have previously been shown to inhibit cell growth and product formation in mammalian cells and to alter the glycosylation pattern of recombinant proteins in cell culture (50). An example of a CO_2-dependent function is the beat frquencey of cilia in airway epithelia, during exhalation the higher concentration of CO_2 present is sensed by sAC which activates PKA and increases ciliary beating (49).

Levels of CO_2 are used as a proxy to detect different things. Insects detect increased CO_2 to locate decaying food and nematode parasites use CO_2 to search for prey (8). Blood-feeding insects, including the malaria mosquito *Anopheles gambiae*, use sensitive olfactory systems to locate hosts. This is accomplished by detecting and following plumes of volatile host emissions including CO_2. CO_2 is

sensed by a population of OSNs in the maxillary palps of mosquitoes and in the antennae of the fruit fly *D. melanogaster* (51). The signals can also be related to internal measures, fungi use high CO_2 levels to induce filamentation and regulate sporulation, and mammals monitor levels to control respiratory exchange (13).

The effect of increased CO_2 levels has been investigated on *C. elegans* as a model system, with the results showing that levels exceeding 9 % cause slowed development and reduced fertility (52). It remains to be resolved whether the receptor binds gaseous CO_2 or HCO_3^-. As CO_2 is an important stimulus for a large number of insect pests, the identification of the CO_2 receptor provides a potential target for the design of inhibitors that might be useful as insect repellents. These would be important weapons in the fight against global infectious disease by reducing the attraction of blood-feeding insects to human hosts (51). Little is known about its role but in *C. elegans*, HCO_3^- regulated adenylate cyclase controls development based on CO_2 levels (13). The avoidance response of *C. elegans* to CO_2 has been shown to be mediated by cGMP (53). There are several guanylyl cyclase enzymes which catalyse the conversion of GTP to cGMP. In mice one of these enzymes, GC-D, has been shown to be activated by the presence of bicarbonate, catalysed from carbonic anhydrase conversion from CO_2 and used in odour recognition (54,55).

This display of CO_2 effects on cellular pH levels, signalling pathways and chemosensing demonstrate a requirement to understand the interactions of CO_2 within a cell and how its effects regulate cellular processes.

1-4 Physiology of CO_2 in plants

1-4.1 Photosynthesis

CO_2 has fundamental roles in plants via its effect on photosynthesis (3) as the substrate for the carbon-fixing enzyme ribulose bisphosphate carboxylase-oxygenase (RuBisCO) to synthesise sugars as well as being a key by-product via biochemical pathways in glycolysis and the Krebs cycle (56).

The Calvin cycle is the name for the light-independent reactions of photosynthesis. The cycle involves carbon fixation, reduction reactions and Ribulose 1,5 bisphosphate regeneration (Figure 1-5).

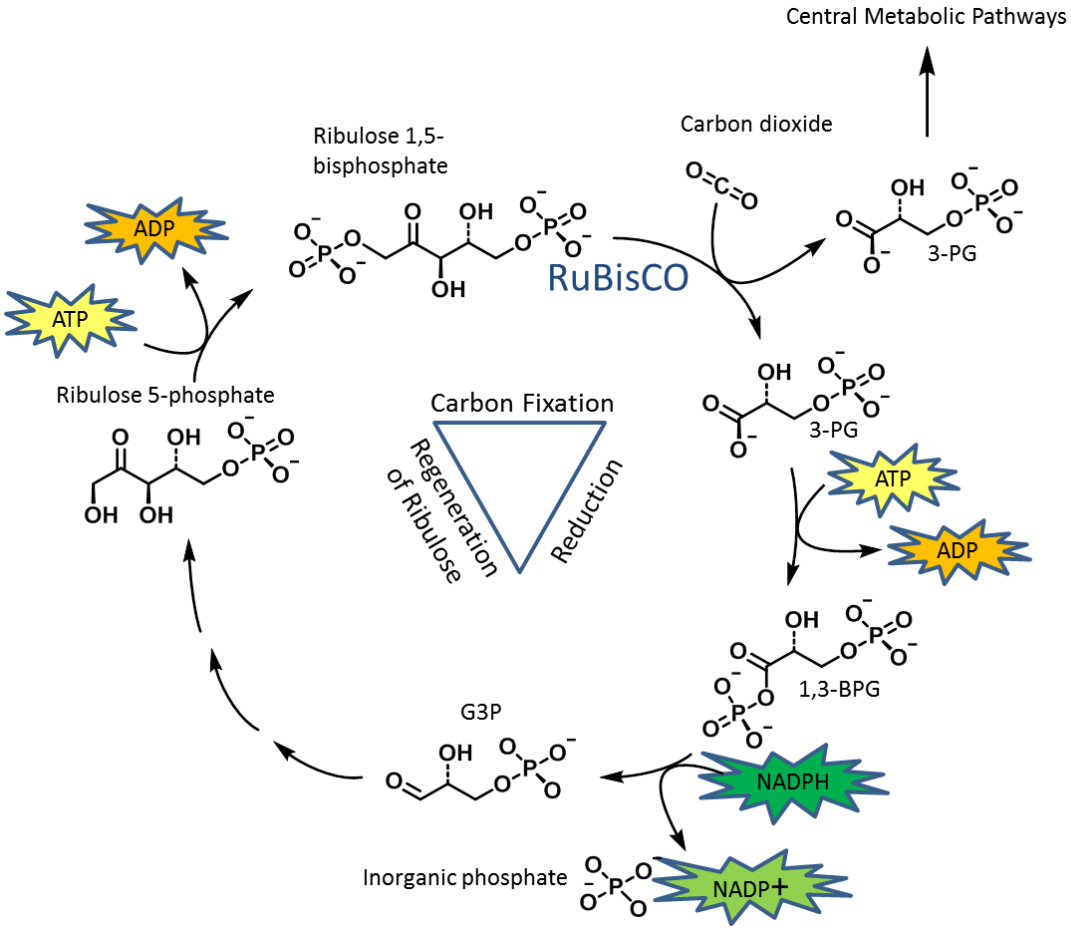


Figure 1-5 The Calvin Cycle showing the role that RuBisCO plays in carbon fixation, the reduction of the 1,3-BPG molecule and the regeneration of RuBP.

Carbon fixation involves the addition of CO_2 to Ribulose 1,5 bisphosphate by RuBisCO. This 6-C molecule is unstable and breaks down to two molecules of 3-PG. One is released while the remaining 3-PG molecule is phosphorylated by phosphoglycerate kinase producing 1,3-BPG that is then reduced by glyceraldehyde 3-phosphate dehydrogenase using NADPH to form G3P. Five molecules of G3P are then used to regenerate three molecules of Ribulose 1,5 bisphosphate (57).

1-4.2 Effect of CO_2 levels on crop yields

Levels of CO_2 are rising (58). There has been a lot of debate surrounding the effect of CO_2 on plants and plant growth, specifically in relation to crop yields. Previous studies of rising CO_2 levels on crop growth carried out in the 1980's (59) suggested that an increase in CO_2 levels would increase crop yields. This was speculated to be due to the RuBisCO photosynthetic reaction occurring in areas that were previously not saturated with CO_2. Many studies have been carried out under controlled conditions involving growth chambers or greenhouses but concern has been voiced that these studies introduce bias surrounding temperature control (60).

It has been demonstrated that these previous studies, all carried out in plant chamber environments, have over-estimated the yield increase by as much as a third. New experiments using a free-air concentration enrichment (FACE) system, in which crops are grown in a more realistic field environment with increased levels of CO_2 pumped over them have shown that in some crops and in tropical growth areas it is expected that crop yields will decrease (61). Studies carried out using the newly developed FACE technique were believed to have overcome most of the previous artefacts but there is still controversy that these new results are exaggerated by the limited available data (62). All these studies demonstrate how difficult it is to accurately predict the reaction of plants to rising CO_2 levels, and highlights the importance of understanding the interaction of CO_2 within a plant system.

All of these studies are limited by the short timescale of the experiments, this means little is known about the effect on future generations of plants grown under these conditions. In one study short term increased atmospheric CO_2 led to an increase in photosynthesis and decreased respiration in terrestrial plants with the C3

photosynthetic pathway which could lead to higher yields. In the longer term this increase was often offset by downregulation of photosynthetic capacity (63).

Another factor to be considered when investigating crop growth is the nutritional value of the crops. The rise of CO_2 levels is most likely to affect the RuBisCO carboxylation reaction, as this reaction occurs in a location not saturated by CO_2. Though most previous studies have focused on the increased growth of plants in increased CO_2 environments (normally by increase in dry weight of material) a study has been carried out to investigate any changes in physiological properties and therefore possible interaction with grazers (64). The study found that though nitrogen efficiency was increased in plants this led to increased phenolics and tannins as well as tissue mass and that the *Chrysophthartus flaveola* beetle did poorly when fed this material in comparison to a matching weight of atmospheric CO_2 grown plants. A second study conducted over 9 years found that plant nutritional quality was reduced by CO_2 increase which led to increased consumption by herbivores to compensate for the reduced nutrition (65). This study even found a reduction in the abundance of herbivores by 21 %, this result suggests that an increase in CO_2 levels could change the ecology of a population.

Understanding how plants will adapt to the rapidly changing climate remains a high scientific priority. The effect on crop yield and nutritional value will have consequences worldwide for both food bases and the stability of ecosystems (66).

Though there is much debate surrounding the effect that increased CO_2 concentration in the atmosphere will have on crop growth it is certain that a difference will occur. The more information that is gathered surrounding the proteins affected by CO_2 within the cell the better prepared we might be for the outcome.

1-4.3 Homeostasis of CO_2 in plants

Homeostasis of CO_2 and pH is also important within plant cells. The *Arabidopsis* gene AtCHX23 has been suggested to encode a putative $Na^+(K^+)/H^+$ antiport within the chloroplast envelope (67). This exchanger is important for regulation of the

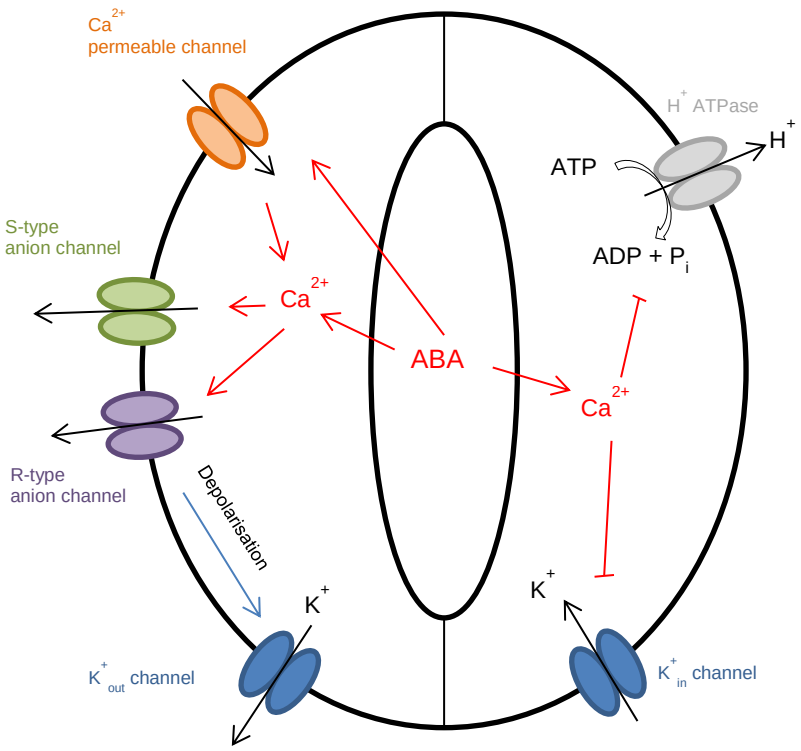
higher stromal pH relative to the chloroplast cytosol. This raised pH is important for regulating several steps of the light-independent photosynthetic enzymes (67).

Besides crop yield there are other mechanisms that could be affected by increased CO_2 levels. For example it has been shown in wheat that an increase in CO_2 level decreases the transcription of RbcS, RbcL and three other Calvin cycle enzymes (68). CO_2 is also transported in plant cells in a similar way to mammalian cells. AQP channels are so important it was speculated that there must be a homologous protein in plants, NtAQP1 was found to transport CO_2 in the same way as mammalian aquaporins (69).

1-5 Detection pathways of CO_2 in plants

CO_2 enters plants through stomata pores on the leaf surface. Stomata pores are made up of a pair of guard cells which form the main entrance point of CO_2 into a plant. Guard cells control the opening and closing of stomatal pores by sensing light, drought and CO_2. The process of CO_2 entrance through this pore is via a complex signalling pathway involving kinases/phosphatases, hormone stimuli and ion channel regulation (70). The stomatal closing of many plant species is dependent on free Ca^{2+} concentration which mediates CO_2 signal transduction in guard cells (71).

Stomatal aperture is regulated by abscisic acid (ABA), ROS, NO and Ca^{2+} ions (72). ABA is the main hormone involved in the opening and closing of the stomata, stomatal closing starts with ABA-induced signalling which increases cytosolic Ca^{2+}. This increased Ca^{2+} inhibits proton and K^+_{in} pumps as well as activating several anion pumps called S and R-type channels. These two steps combine to cause an anion efflux within the cell which causes depolarisation. This depolarisation additionally activates K^+_{out} channels which contribute to the anion efflux and the guard cell loses its turgor causing stomata closure (73) (Figure 1-6).



Stomata Closing

Figure 1-6 Simplified pathways involved in the process of stomata closing and opening beginning with ABA signalling.

Reversible protein phosphorylation is known to play a crucial role in guard cell control of stomatal opening (74). The opening of stomata to allow CO_2 diffusion needs to carefully balance with the loss of water vapour out of the plant. Opening of stomata through guard cell swelling is caused by the uptake of K^+ ions.

CO_2 is present in all organisms that undergo cellular respiration (75) and is monitored by K^+ channels in all three domains of life (76). K^+ channels in guard cells are complex and regulated by several signal molecules, one of which is CO_2 (77). Work was carried out to investigate the response to K^+ channels in *Vicia faba* in response to increased CO_2 levels. Though the mechanism is unknown it was found that the activity of the K^+ channels was reduced by increased CO_2 independent of pH (77).

Another study on the regulation of K^+ channels concluded that several factors were involved in the regulation of the channels, including CO_2 and pH but that how the factors are linked to the channel function has yet to be discovered (76). K^+ channel regulation by pH is widely known and is implicated in the use of CO_2 during chemosensing (76).

Plants depend on CO_2 entering through open stomata for photosynthesis. Further knowledge into the specific mechanism of stomata opening to CO_2 enrichment would provide a lot of information into the effect of increased CO_2 levels (78). Though the consequences of increased CO_2 remain unknown it is certain that the increase will affect stomata as they are the entrance for CO_2 into the cell.

Homeostasis in plants surrounds the opening and closing of stomata to regulate the entrance of CO_2 into leaf cells. The discovery of SLAC1, a slow anion channel, has been found to mediate CO_2 sensitivity in the regulation of plant gas exchange. The efflux of Cl^- and $malate^{2-}$ ions was already known to occur in guard cells but SLAC1 is the first evidence of the anion channel carrying out this ion exchange. SLAC1 is localised to the plasma membrane of guard cells and its activation increases with the rise of CO_2 (79). This process occurs by the regulation of anion transfer within guard cells which controls the turgor of the cells.

1-6 Carbamylation the post-translational modification

The pathways of CO_2 that have been discussed are all identified by observed system alterations with changing CO_2 levels. How CO_2 is physically interacting within the pathway is often unknown. One way in which CO_2 can interact within a protein system is via the formation of a carbamate on an amine group.

1-6.1 Types of Carbamylation

There has been some confusion surrounding carbamates in the literature as carbamylation is a term that has been used to describe two different PTMs. One of which is the reaction with CO_2 which is discussed in this work. The other is the interaction of an amine group with urea, or a urea break-down product isocyanate. The latter is not a labile reaction and has been analysed using mass spectrometry techniques (80) (Figure 1-7).

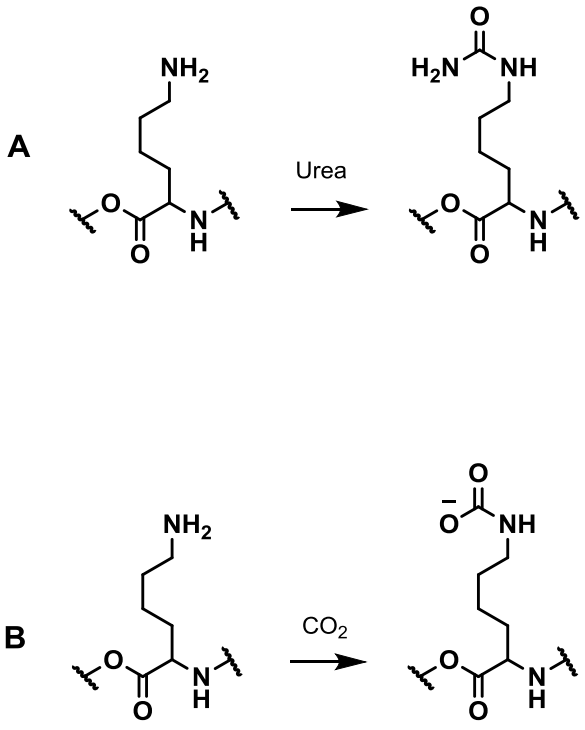


Figure 1-7 Schematic difference between the isocyanate formed carbamate (A) and the CO_2 carbamate (B).

1-6.2 Carbamate formation and equilibria

The carbamate formation important for this work involves the interaction of CO_2 with amines. CO_2 is generally unreactive but it combines rapidly with amines at cellular temperatures and pressures to form carbamates. This reaction is the nucleophilic attack of neutral amines on CO_2 through a reversible acid-base reaction (81). Carbamylation is one of the earliest post-translational modifications (PTMs) to be identified (82). In a cellular system carbamylation converts side chain lysine or N-terminal residues, from basic/neutral to acidic functionality (Scheme 1-4) (83).

Scheme 1-4 The reversible nucleophilic attack of a neutral amine on CO_2 to form a carbamate.

The formation of a carbamate can alter peptide structure or binding which may lead to an effect within a cellular system (84). Carbamylation is under investigated due to the labile nature of the modification making it difficult to study (83-85). As a reversible process carbamates were previously thought to be of little biological significance.

It has since been suggested that carbamylation may be more widespread than previously realised and not limited to the proteins so far recognised (86). The formation of a carbamate links to the equlibria of CO_2 in solution. In order for a carbamate to form, the CO_2 needs to be in aqueous form not as bicarbonate or carbonic acid.

There has been debate over the formation of a carbamate and the conditions that are needed for such an interaction, it has been observed that an increase in pH increases the amount of carbamate formed (87). This implies that the rate limiting step of the carbamate formation is the charge on the amine group and is linked to the pKa. The pKa of an amine group on a lysine residue ranges from 9-13 depending on the specific environment (88). The pH of a physiological environment

is around 7.4 implying that carbamate formation is enhanced by protein structure and the formation of restricted hydration spaces to encourage reduced lysine pKas.

As can be seen from Figure 1-8 the formation of a carbamate involves the amine group being in its uncharged state as well as the conversion of CO_2 into its aqeuous form from bicarbonate. The pKa of the average lysine side chain is 9-10. This means that under normal physiological conditions of pH 7.4 most amine groups are in a charged state. It is therefore suggested that carbamates are formed on 'privledged' amines.

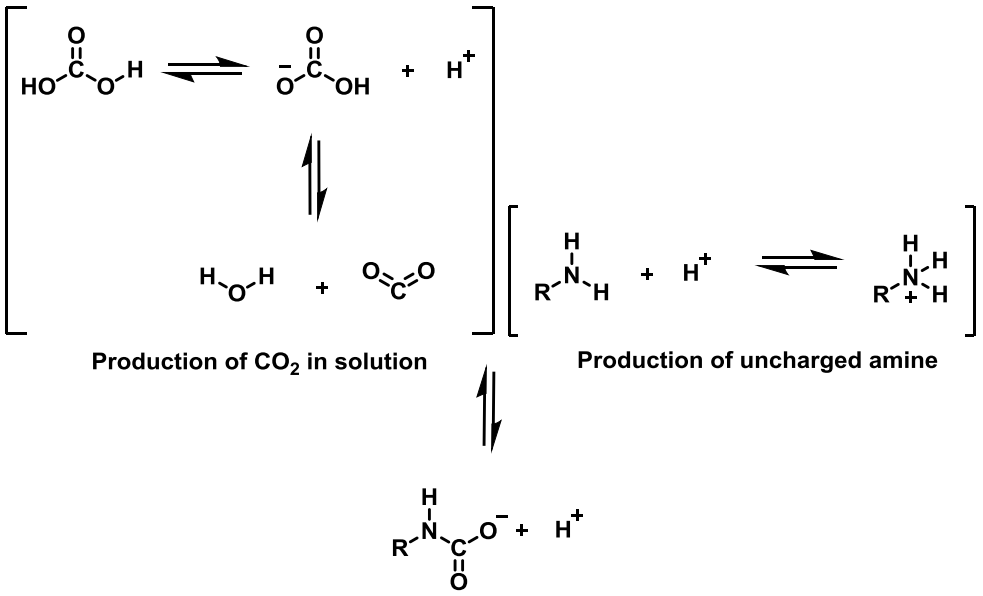


Figure 1-8 Demonstrating the combination of the equilibria involved in the formation of CO_2 and an uncharged amine that are both necessary for carbamate formation.

A number of proteins have been found to interact with CO_2 by the formation of a carbamate on a privileged NH_2 group; these are discussed in more detail in 1-7, 1-8 and 1-9.

When investigating a carbamate on a single amino acid or short peptide system the priviledged environment of a protein structure will not be present. *Le Chateleirs* principle could help with the formation of possible carbamate to trap. The principle suggests that a change to a system in equilibrium will cause an adjustment to

counteract this change as far as possible. This implies that as the uncharged amines form a carbamate more amine will become uncharged (89).

It is also suggested here that the fold of a protein can produce an environment capable of encouraging carbamate formation. For example the *N*-terminal valines on β-haemoglobin chains are contained within such an environment to encourage the carbamate formation having a pKa of 6.6 (90).

1-7 Carbamylated proteins in mammals

1-7.1 Haemoglobin

There are few confirmed carbamylation sites because of the modifications labile nature. The first of these is the carbamino group formation of haemoglobin (Hb) which alters its oxygen binding properties with the binding of CO_2 (84). An increase in blood CO_2 concentration leads to a decrease in blood pH which results in Hb releasing the bound oxygen; this is known as the Bohr effect (91). Within a capillary a carbamate is formed at the *N*-terminal residue of each of the Hb chains producing carbaminohaemoglobin (91). Previous studies of this modification have found that the adduct formed on the β-chain are more prominent than on the α-chain (92). This binding of CO_2 to Hb causes a stabilising effect on the protein which changes its oxygen binding properties (93). 10-20 % of CO_2 transport in the blood is carried by Hb in this way (93). Deoxygenation of the blood increases its ability to carry CO_2 i.e. the lower the saturation of Hb with O_2 the larger the CO_2 saturation, this is known as the Haldane effect (94).

The presence of carbamylation adducts in Hb was originally confirmed using ^{13}C NMR spectroscopy (95). This technique could not identify the location of the carbamate, only that a carbamate was forming on the protein. The site of carbamate formation was later recognised by mutating the *N*-terminal residues with cyanate and investigating the formation of deoxyhaemoglobin. This technique of mutation is unsuitable for identifying the location of many carbamates. This is because they are often located in active sites and their mutation would therefore render the protein not functional.

1-7.2 Connexin 26

Gap junctions are composed of two hemichannels, these channels comprise primarily of connexin proteins (96). Connexin 26 (Cx26) has been specifically shown to be sensitive to P_{CO2} levels by the formation of a carbamate on Lys125. An experiment carried out by Meigh *et. al.* 2013 demonstrated channel opening to

increased CO_2 levels despite a constant pH (97). The effect of carbamylation was tested by mutation of the Lys125 residue. Previously the main theory of P_{CO2} measurement has been via pH-sensing and only recently it has been suggested that direct CO_2 or HCO_3^- molecules are being detected (98).

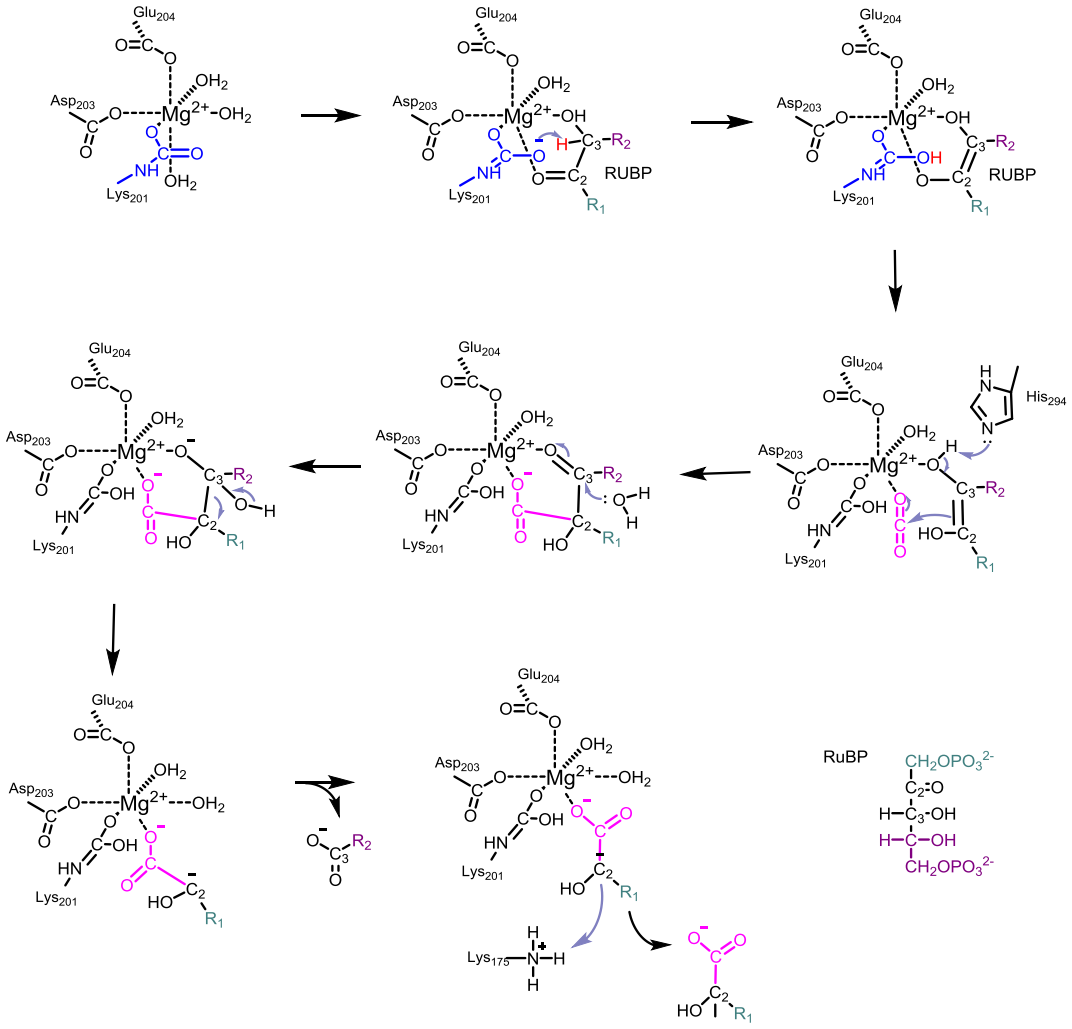
It has been shown that three hemichannels (Cx 26, 30 and 32) can be opened by CO_2. These channels all contain a carbamylation motif discovered only in the CO_2 sensitive connexions which suggest that this CO_2 regulation is through the formation of a carbamate (99).

1-8 Carbamylated proteins in plants

1-8.1 RuBisCO

As mentioned previously the major route of CO_2 metabolism in the environment is the Calvin-Benson-Bassham (CBB) reductive pentose phosphate pathway, the main enzyme in this pathway is RuBisCO (1). RuBisCO is one of the most abundant enzymes on Earth (100), it catalytically fixes CO_2 in photosynthesis, on the order of 10^{11} tons each year (2,101). Being such a large part of photosynthesis RuBisCO is present in most autotrophic organisms including algae and plants (102) and has been said to represent the ultimate energy source due to its linking inorganic and organic carbon (103,104).

It has been shown that activation of the RuBisCO for its carboxylase function requires the stepwise addition of CO_2 binding through formation of a carbamate followed by stabilisation by a magnesium ion, with the CO_2 addition being the rate limiting-step (83). This activator CO_2 is different from the later CO_2 which becomes fixed during carboxylation (105) (Scheme 1-5).



Scheme 1-5 The process of Lys201 carbamate formation within the RuBisCO active site catalysing the conversion of RuBP to two 3-PG molecules, one is released and one is used in the regeneration of RuBP.

RuBisCO represents the first enzyme known to be activated by CO_2 through carbamate formation, as well as using it as a substrate (105-107). Both of the RuBisCO functions require carbamylation of an internal lysine residue (104).

NMR spectroscopy studies with $^{13}CO_2$ revealed the presence of binding of CO_2 to RuBisCO by the formation of a carbamate. The site of the carbamate (Lys201) was established by X-ray crystallography. This technique involves higher than physiological concentrations of CO_2 and non-physiological solvent conditions which could cause artefacts. Therefore this method is not applicable for searching for unknown carbamates under cellular conditions.

1-8.2 Urease

Urease is an enzyme with an important role in the metabolism of nitrogen in plants and microbes by catalysing the hydrolysis of urea to form ammonia and carbon dioxide. It has been isolated from various plants, fungi and bacteria. The urease active site contains a carbamylated lysine (Lys217) residue stabilised by a nickel atom (108). The enzyme has been shown to be CO_2 dependent in vitro which is expected to be due to the presence of the carbamylated lysine (109). The site of carbamate formation was identified using X-ray crystallography.

The investigation by Pearson *et. al.* 1998, was interested in whether the carbamate chemistry was essential for activation of urease or if it could be substituted by a carboxylate group. Active site mutations of urease replaced the carbamylated lysine with a glutamate to look at interaction with the nickel metal ions. No activity was achieved as it was seen that the glutamate residue was not long enough to reach the metal binding site (109). The activity was later rescued by the addition of formic acid which was able to bridge the nickel centre. This implies that the chemistry introduced by the carbamylated lysine is essential for enzyme activity.

1-9 Carbamylated proteins in bacteria

1-9.1 Alanine Racemase

The carbamate modification is also present in bacteria. Alanine racemase is the enzyme responsible for catalysing the formation of D-alanine which is used by bacteria to synthesise peptidoglycan for cell walls (110). A step in the process of this catalysis involves the formation of a carbamate of the side chain of Lys129 which is thought to influence the substrate binding and catalysis by correctly positioning the Arg136 residue (110). The absolute requirement for D-alanine in peptidoglycan biosynthesis makes alanine racemase an attractive target for inhibitors that might also function as antibiotics (110).

In the case of alanine racemase, the close proximity of Arg136 to Lys129 would be expected to promote deprotonation of the lysine. Arg136 is also perfectly positioned to stabilise the carbamate product once formed similar to how a metal ion stabilises carbamates in other enzymes with lysine carbamylation.

This carbamate location was discovered using electron density mapping. Much like the crystallisation carried out for RuBisCO, these conditions do not match a physiological environment so could introduce artefacts.

1-9.2 Transcarboxylase

Transcarboxylase Biotin is a CO_2 cofactor that carries CO_2 in the form of carboxybiotin. Transcarboxylase from *Propionibacterium shermanii* is a biotin-dependent carboxylase, it is involved in the transfer of CO_2 to pyruvate using two reactions (111). The first of which transfers CO_2^- from methylmalonyl-CoA to biotin and the next moves that same CO_2^- to pyruvate (112). The large transcarboxylase subunit 5S is homologous to the C-terminal carboxyltransferase region of human pyruvate carboxylase, therefore work on this transcarboxylase enzyme can elucidate information about the mammalian enzyme (112).

A carbamylated lysine residue has been discovered within the active site of this subunit coordinated to a cobalt ion. As well as coordinating this cobalt the carbamylation modification also forms hydrogen bonds with several residues all of which appear to be conserved and involved in the catalysis of carboxylating biotin (113). It has been shown that catalysis fails without the presence of the carbamylated lysine (112). This structural information was gained by X-ray crystallography.

1-9.3 β- lactamase

β-lactam antibiotics prevent cell wall synthesis in bacteria and are therefore a line of defence in the treatment of bacterial infections (114). These antibiotics are challenged by β-lactamase resistance enzymes. These enzymes come in four classes, A, B, C and D. Class D are currently the least studied but have been shown to form a carbamate for their catalytic abilities (115). The crystal structure of OXA10 class D β-lactamase from *Pseudomonas aeruginosa* has been solved revealing many newly discovered features of the protein structure but most notably the formation of a carbamate on the active site residue Lys70 (114).

In the previously mentioned enzymes the carbamate formation is stabilised by interaction with a metal cation. The formation of a carbamate at Lys70 on β-lactamase is the only known example of an active site lysine carbamylation which does not coordinate to a metal ion. Instead it is stabilised by hydrogen bond interactions to other active site residues and a water molecule (114).

Work on β-lactamases utilised labelled bicarbonate (e.g. $NaH^{13}CO_3$) and analysis by ^{13}C NMR spectroscopy to investigate the carbamylated residue (115). This site was also confirmed by crystallography which puts the enzyme under non physiological conditions. The study also examined the activity of the enzyme with regards to carbamate formation. It was found that the enzyme was inactive when in degassed buffer but regained activity with the addition of bicarbonate at pH 7.0 (114). The results of this assay demonstrate the importance of carbamate formation in enzyme active sites.

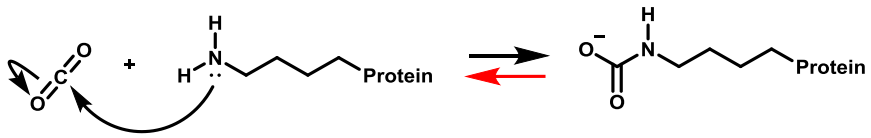
1-9.4 Phosphotriesterase

Phosphotriesters are insecticides that are used in the protection of crops by working as nerve agents (116). Bacterial phosphotriesterases evolved to be able to hydrolyse these compounds after their introduction into the environment. These enzymes detoxify pesticides and require a binuclear metal centre for activity (117).

The reaction mechanism of these enzymes is thought to proceed through an S_N2-type catalytic mechanism (118). The active site contains two metal ions, though it is not clear which metal ion is present, but one of the bridging ligands is a carbamate formed on Lys169. The enzyme was assayed for activity prior to crystallisation but in some cases the metals were not present in the active site due to the conditions required for crystal formation. This gives an example of how differences can be caused by the artificial conditions required for crystallisation.

1-10 Practical challenge

The main reason for the lack of research to investigate carbamate formation is its labile nature. This means that the modification freely dissociates as seen by the red arrow in Scheme 1-6.

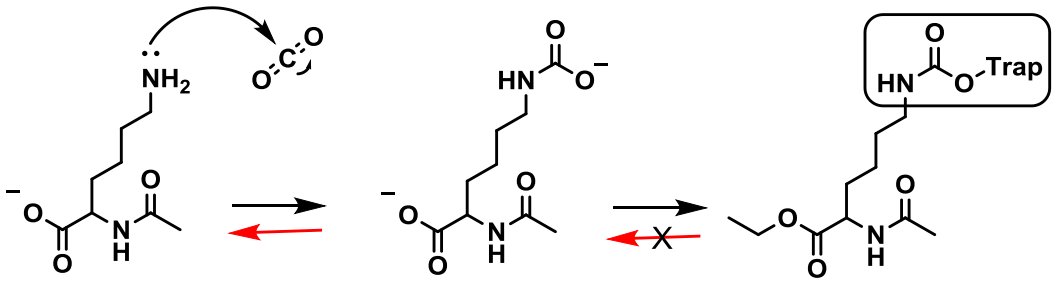


Scheme 1-6 Labile reaction mechanism of carbamate formation.

The main method of PTM analysis is MS. Previously it was not possible to analyse many PTMs in this way due to the necessity of hard ionisation techniques. In recent years soft ionisation techniques such as MALDI and ESI have made it possible to analyse the formation of a carbamate on a whole protein (119). However, the additional energy required to fragment the protein for site confirmation causes the release of the carbamate modification and therefore cannot identify the location.

There are several previous investigations into the formation of carbamates on proteins. The main difference between the aims of this work and previous studies is to perform the fixing of the CO_2 under conditions closely matching a cellular environment and then allow analysis of the site. Previous studies looking at carbamate site location have been performed under organic solvent systems. This could mean that their results are artificially formed under the experimental conditions (120). Several studies have confirmed the presence of a carbamate in a protein through crystallisation but this again cannot be confirmed as a natural state carbamate (106).

To overcome this challenge a trapping method will be developed. This work aims to remove the labile nature of the carbamate and create a more robust molecule for use in MS. The hypothesis here is that the transfer of a chemical group to the formed carbamate will provide this robust nature and prevent the dissociation reaction (Scheme 1-7).



Scheme 1-7 Converting a carbamate PTM into a robust modification for downstream analysis by removing the labile nature.

The group transfer needs to occur under physiological conditions of pH and CO_2 concentration in an aqueous environment to provide a system that can more accurately predict the location of a naturally occurring carbamate without artefact. This technique will also provide a method that can be used on any protein system, some proteins are difficult to crystallise so would not be applicable for the current discovery techniques.

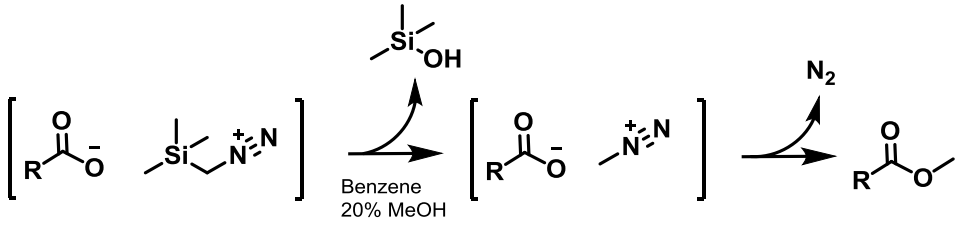
1-11 Previous trapping work

Thorough detail of the previous background related to trapping of carbamates is included within the results chapters 3 and 4 but there is also a brief summary here.

MS has been used to identify carbamates previously, however due to the fragility of the adduct being investigated, the location of the modification cannot be identified (84). A study was carried out to investigate the soft ionisation techniques that exist for MS to look at their ability to visualise a carbamate. Terrier and Douglas (2010) were able to identify the formation of a carbamate at high pH and using dissolved CO_2. Unfortunately these modifications were mostly on peptides and due to the high pH values that are not physiologically relevant, likely to be artificial.

1-11.1 TMS-DAM

TMS-DAM is an electrophilic methylating agent and has been used previously in organic systems for the *O*-methylation of alcohols in dichloromethane (121). Work using TMS-DAM to trap carbamates has also been carried out previously, but only within a mostly organic (benzene/methanol (4:1)) solvent environment (122) (Scheme 1-8).



Scheme 1-8 Esterification of a generic carbamate by TMS-DAM in an organic solvent system.

This reaction occurs by the transfer of a methyl group from TMS-DAM to the carbamate. This reagent was able to successfully esterify the carbamate but the organic solvents cannot be used in this work because any reaction under non-physiological conditions could create artificial results. Therefore one proposal in this synthetic work was to produce a water-soluble derivative of TMS-DAM that can be

used to esterify carbamates in the same way but within an aqueous system. (The results describing this work are contained in Chapter 3).

1-11.2 Meerwein's reagents

A second method to be investigated was the esterification by an electrophilic reagent, specifically Meerwein's reagent. This term covers both the trimethyloxonium (TMO) and triethyloxonium (TEO) tetrafluoroborate salts (Figure 1-9).

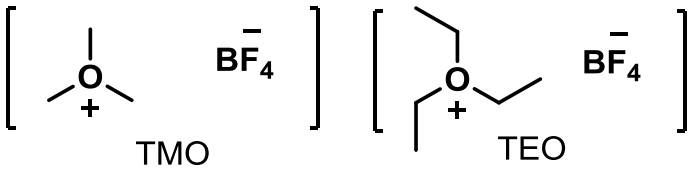
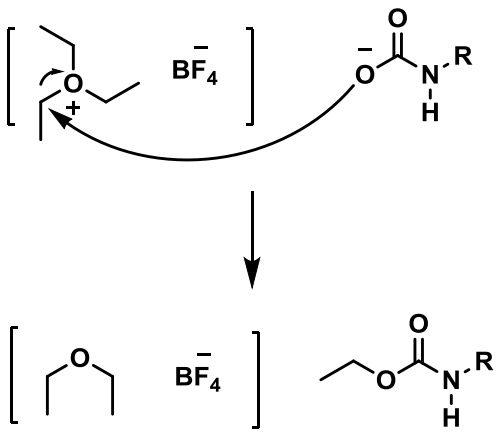


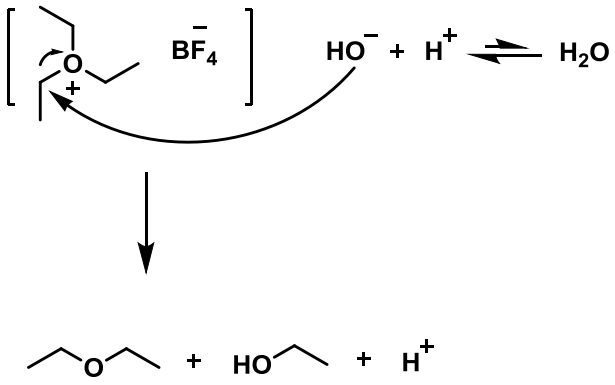
Figure 1-9 Meerwein's reagents TMO and TEO.

The trapping process of Meerwein's reagent occurs via a similar mechanism as described for TMS-DAM, with the transfer of a methyl or ethyl group to the CO_2 bound to an amine residue (Scheme 1-9).



Scheme 1-9 Hypothesised ethylation of a carbamate by TEO.

Both TMS-DAM and Meerwein's reagents hydrolyse in water, with the hydroxide ion acting as the nucleophile (Scheme 1-10).



Scheme 1-10 Hydrolysis of Meerwein's reagent TEO to form ethanol, ether and H+ ions.

This is likely to be the reason why these reagents have not often been used for reactions in aqueous solution. The Meerwein reagents have an advantage over diazomethane, with triethyloxonium having a half-life in water on the order of 7 minutes (123) which is far greater than the 1.8 seconds found for diazomethane in a THF/water mix (124). This gives the reagent longer before hydrolysis to trap the desired carbamate. Trimethyloxonium has also been suggested to have a selectivity for a carboxyl group over a water hydroxide group (125).

Some work using these reagents in aqueous solution to investigate the modification of carboxyl groups have countered this hydrolysis by the use of a large excess of reagent (123). Though one of the reagent side products is ethanol this is not expected to cause a denaturation problem within the reaction conditions as the maximum concentration of ethanol produced accounts for 0.25 % volume of the reaction and a concentration of 22 % volume is needed to see half of a protein mixture denatured (126). Previous work using Meerwein's reagents to investigate carboxyl group modifications and the development of their use to trap carbamates is described in detail in Chapter 4.

1-12 Motivation for investigation

Reactions of CO_2 with amino groups have been shown to markedly influence the function of several proteins including two highly abundant proteins; Hb and RuBisCO. It is therefore likely that other carbamylation sites alter functions of other proteins.

Carbamates are understudied due to their labile nature. Analysis of most PTMs is carried out by MS but the fragility of the carbamate modification limits its downstream processing. This has led to most carbamate identification being performed under conditions that do not match a cellular environment, such as required for x-ray crystallography.

Recently Jimenez-Morales *et. al.* 2014 have developed a computational method to predict the formation of carbamates on lysine residues. The results of this study suggest that as many as 1.3 % of proteins may have a carbamylated residue (85).

1-12.1 Project relevance

It is remarkable that we know so little of how CO_2 influences the function of the proteome. The main reason for this is that the interaction of CO_2 with proteins is labile and difficult to study without large amounts of purified protein and using artificial conditions.

The binding interactions of CO_2 to proteins have been previously investigated to be utilised for carbon capture and sequestration. Current technologies in these areas are often corrosive or produce toxic waste and greener technologies involving proteins are being sought (127).

There is also a clinical relevance to understanding deeper CO_2 mechanisms. Under normal circumstances the body is equipped to buffer changes in levels of inorganic carbon using homeostasis (97), however when an underlying disease disrupts this it can lead to a pathological acid-base disorder. It can affect tissues and body fluids and is thought to have influences on respiration and modes of metabolism (128).

Normal levels are essential for the function of many tissues and abnormal levels can lead to hyper- or hypocapnia (increased or decreased blood CO_2 concentration) and effect calcification (129). Hypercapnia has been reported to cause pathophysiological effects in all tested eukaryotes (75). These effects are thought to occur due to cells being able to sense CO_2 levels and alter their growth (42) with suppression effects seen on NF-$_K$β. NF-$_K$β is one of the most understood transcription factors and regulates expression of many genes. Importantly it is required in the encoding of proteins involved in immunity and inflammation (130). The specific molecular mechanisms of these pathways remain unclear.

Atmospheric CO_2 levels are still rising and this increase is predicted to have a significant impact on crops (131). It has been found that decreased water uptake due to elevated CO_2 could be detrimental to the hydrological cycle (132). Elevated CO_2 has also been found to decrease levels of zinc and iron as well as protein levels in several crop plants (133). The true impact of increased CO_2 levels on plants is still widely speculative due to the small sample size of most experiments. Knowledge of the interactions between CO_2 and proteins is therefore vital for understanding more about the effects the rise in CO_2 could have on crop plants. It has been seen that giving plants increased CO_2 increases the rate of photosynthesis for a short period of time but it is not sustained (134). If more information were known about the process it is possible the proteins involved could be exploited to increase carbon fixation.

The true extent of signalling involving CO_2 is not even close to being known as so many enzymes utilise CO_2 during activity (42). Increasingly carbamates are being seen in important protein systems. There is much debate over the link between acid-base balance in cells and pH/CO_2 concentration. Most believe that it is pH-sensing which causes effects seen by molecular CO_2, however recent work carried out looking at connexin hemichannels has looked specifically at effects seen when CO_2 concentration is increased but pH is kept constant. This demonstrated the channel to be effected by the change in CO_2 independent of pH. Carbamylation is a possible way that CO_2 could be a signalling molecule directly interacting with proteins, this has been found to be the case for the protein connexin 26 (97).

The identities of plant proteins which interact directly with CO_2 are almost entirely unknown. Therefore it is essential to develop tools to identify CO_2 targets to improve the understanding of plant CO_2 interactions. This work was specifically aimed at CO_2 binding upon plant leaf proteins. However the importance of understanding CO_2 to protein binding is prevalent in many areas.

1-12.2 Aims and Hypothesis

We hypothesise that there are far more sites of CO_2 interaction upon proteins than has ever been previously anticipated. We also hypothesise that the main form of CO_2 interaction within a protein system is through the formation of a carbamate.

This investigation has two main aims. The first of which is to develop a method to convert the labile modification of a carbamate into a more robust molecule that can be analysed by downstream mass spectrometry methods. This method needs to be applicable to be used under conditions which accurately represent the physiological environment of a cell.

The second aim following from this is to use this method to trap a carbamate on an unknown protein within a cellular system and discover the site of this trapping to provide new information about the carbamates that are formed.

The work in this thesis describes two approaches towards these goals: 1) the chemical synthesis of a water-soluble TMS-DAM derivative (chapter 3) and 2) the direct use of Meerwein's regents TMO and TEO (chapter 4). Both methods of trapping are based on the transfer of a methyl or ethyl group to the carbamate to create a more robust PTM for downstream analysis.

The development of a trapping method to achieve these aims will begin with an amino acid system, then progress to a dipeptide, then to purified proteins known to form a carbamate such as Hb and RuBisCO, and finally on to other proteins working within CO_2 environments. This trapping method will then be applied to a soluble leaf lysate from *Arabidopsis* to identify new sites of carbamylation (Chapter 5).

Chapter 2: Materials and Methods

2-1 Materials and Equipment

All materials were purchased from Sigma-Aldrich unless otherwise stated. NMR spectra were recorded on a Bruker 400 spectrometer with all J-coupling values given in Hz. ESI Mass spectra were obtained on a Waters Ltd QToF or Thermo LTQ XL Orbitrap. Samples were lyophilised on an LP3 Jouan high vacuum system. The pH measurements were carried out on a pH stat Radiometer analytical TIM856, the probe was calibrated using standards at pH 7 and 10 (Fisher).

2-1.1 Cell lines

E. coli DH5α (no antibiotic resistance) cells were used for all DNA manipulation work and *E. coli* BL21 (Novagen, no antibiotic resistance), Rosetta and Rosetta 2 (Novagen, both chloramphenicol resistant when containing the pLysS plasmid) cell lines were used for protein expression.

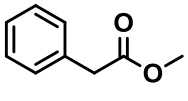
2-1.2 pH stat

All pH stat experiments were carried out in phosphate buffer (4 mL, 50 mM, pH 7.4). This solution was transferred to the pH stat and incubated at 25 °C with stirring. Triethyloxonium tetrafluoroborate (various amounts) was added stepwise with a constant pH maintained (pH 7.4) with the slow addition of 1 M NaOH. The reaction was stirred for 1 h.

2-2 Experimental Biology

2-2.1 Trimethyloxonium tetrafluoroborate trapping experiments

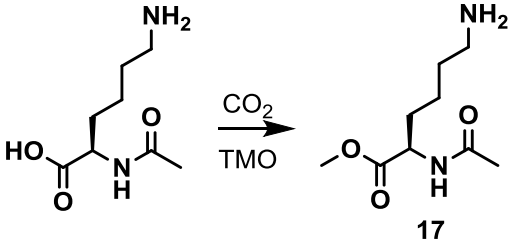
2-2.1.1 Phenylacetate



16

Phenylacetic acid (4.4 mg, 0.03 mmol) was dissolved in phosphate buffer (2 mL, 50 mM, pH 7.4) and trimethyloxonium tetrafluoroborate (TMO) (130 mg, 0.88 mmol) was added at room temperature. The reaction mixture was stirred for 1.5 h and the product extracted with diethyl ether (2 × 5 mL). The ethereal layers were combined and the solvent was removed under pressure to yield phenylacetate **16** (4.1 mg, 93%). ^{1}H NMR (400 MHz, $CDCl_3$) δ/ppm 7.37-7.27 (5H, m, Ar-H), 3.72 (3H, s, CH_3), 3.66 (2H, s, CH_2). ^{13}C NMR (100 MHz, $CDCl_3$) δ/ppm 171.9, 133.9, 129.2, 128.5, 127.0, 51.9, 41.1. GC-MS [M+H] 150.08.

2-2.1.2 Acetyl-Lysine and TMO experiment

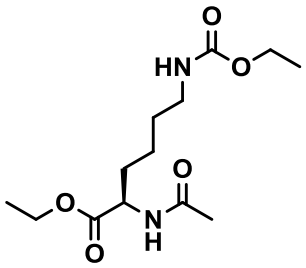


17

Acetyl-lysine (10 mg, 0.05 mmol) was dissolved in phosphate buffer (2 mL, 200 mM, pH 7.4). Sodium bicarbonate ($NaHCO_3$) (1.7 mg, 0.02 mmol) was dissolved in phosphate buffer (1 mL, 200 mM, pH 7.4) and added to the acetyl-lysine. TMO (50 mg, 0.34 mmol) was added with stirring, while the pH was maintained by manual addition of sodium hydroxide (NaOH) (1 M, Fisher Scientific), the reaction was stirred for 1 h. ESI-MS showed methyl transfer on the carboxyl group (**17**) but no carbamate formation. ESI-MS [M+H] 203.14

2-2.2 Triethyloxonium tetrafluoroborate trapping experiments

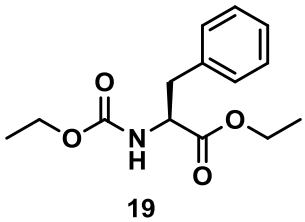
2-2.2.1 Acetyl-Lysine trapping



18

Acetyl-lysine (5 mg, 0.03 mmol) was dissolved in phosphate buffer (2 mL, 50 mM, pH 7.4). $NaHCO_3$ (1.7 mg, 0.02 mmol) was dissolved in phosphate buffer (1 mL, 50 mM, pH 7.4) and added to the acetyl-lysine solution. This solution was transferred to the pH stat. Triethyloxonium tetrafluoroborate (TEO) (100 mg, 0.53 mmol) was then added to the reaction in three portion-wise steps while the pH of the solution was maintained with automated addition of NaOH (1 M). The reaction mixture was stirred for 1 h then lyophilised and re-dissolved in methanol (1 mg/mL) for MS. The sample was analysed using ESI-MS and the trapped carbamate acetyl-lysine product **18** was confirmed. ESI-MS [M+H] 289.17.

2-2.2.2 Phenylalanine trapping

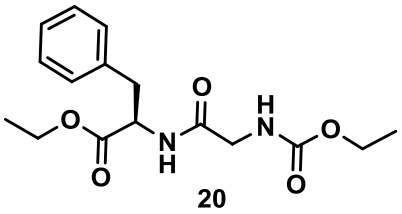


19

Phenylalanine (5 mg, 0.03 mmol) was dissolved in phosphate buffer (2 mL, 50 mM, pH 7.4). $NaHCO_3$ (1.7 mg, 0.02 mmol) was dissolved in phosphate buffer (1 mL, 50 mM, pH 7.4) and added to the phenylalanine solution. This solution was transferred to the pH stat. TEO (100 mg, 0.53 mmol) was then added to the reaction in three portion-wise steps while the pH of the solution was maintained with automated addition of NaOH (1 M). The reaction mixture was stirred for 1 h then lyophilised and

re-dissolved in methanol (1 mg/mL) for MS. The sample was analysed using ESI-MS and the trapped carbamate phenylalanine product **19** was confirmed. ESI-MS [M+H] 266.13.

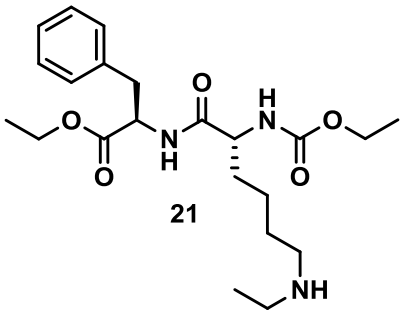
2-2.2.3 GLY-PHE dipeptide trapping



20

GLY-PHE dipeptide (8 mg, 0.04 mmol) was dissolved in phosphate buffer (2 mL, 50 mM, pH 7.4). $NaHCO_3$ (1.7 mg, 0.02 mmol) was dissolved in phosphate buffer (1 mL, 50 mM, pH 7.4) and added to the dipeptide solution. This solution was transferred to the pH stat. TEO (280 mg, 1.47 mmol) was added to the reaction mixture in three portion-wise steps in dH_2O (1 mL) while the pH was maintained with automated addition of NaOH (1 M). The reaction mixture was stirred for 1 h then lyophilised and re-dissolved in methanol (1 mg/mL) for MS. The sample was analysed using ESI-MS and the trapped carbamate GLY-PHE product **20** was confirmed. ESI-MS [M+H] 323.01.

2-2.2.4 LYS-PHE dipeptide trapping

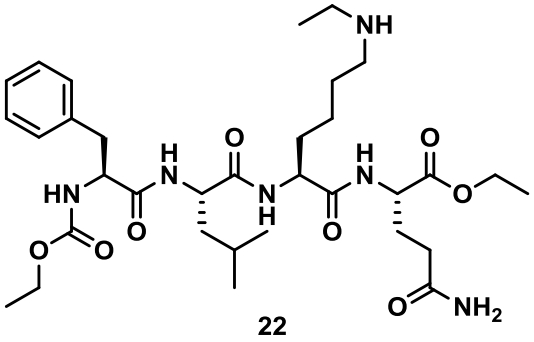


21

LYS-PHE dipeptide (8 mg, 0.04 mmol) was dissolved in phosphate buffer (2 mL, 50 mM, pH 7.4). $NaHCO_3$ (1.7 mg, 0.02 mmol) was dissolved in phosphate buffer (1 mL, 50 mM, pH 7.4) and added to the dipeptide solution. This solution was added to

the pH stat. TEO (280 mg, 1.47 mmol) was added to the reaction mixture in three portion-wise steps in dH_2O (1 mL) while the pH was maintained with automated addition of NaOH (1 M). The reaction mixture was stirred for 1 h then lyophilised and re-dissolved in methanol (1 mg/mL) for MS. The sample was analysed using ESI-MS and the trapped carbamate LYS-PHE product **21** was confirmed. ESI-MS [M+H] 422.18.

2-2.2.5 FLKQ tetrapeptide trapping



22

FLKQ tetrapeptide (synthesised by Ashai Cano) (5 mg, 0.009 mmol) was dissolved in phosphate buffer (2 mL, 50 mM, pH 7.4). $NaHCO_3$ (1.7 mg, 0.02 mmol) was dissolved in phosphate buffer (1 mL, 50 mM, pH 7.4) and added to the dipeptide solution. This solution was added to the pH stat. TEO (280 mg, 1.47 mmol) was added to the reaction mixture in three portion-wise steps in dH_2O (1 mL) while the pH was maintained with automated addition of NaOH (1 M). The reaction mixture was stirred for 1 h then lyophilised and re-dissolved in methanol (1 mg/mL) for MS. The sample was then analysed using ESI-MS and the trapped carbamate FLKQ product **22** was confirmed. ESI-MS [M+H] 663.16.

2-2.2.6 Haemoglobin trapping

Human haemoglobin (Hb) (14.5 mg, 0.23 µmol) was dissolved in phosphate buffer (2 mL, 50 mM, pH 7.4). $NaHCO_3$ (1.7 mg, 0.02 mmol) was dissolved in phosphate buffer (1 mL, 50 mM, pH 7.4) and added to the protein solution. This solution was added to the pH stat. TEO (280 mg, 1.47 mmol) was added to the reaction mixture

in three portion-wise steps in dH_2O (1 mL) while the pH was maintained with automated addition of NaOH (1 M). The reaction mixture was stirred for 1 h then dialysed overnight (1 L dH_2O). The sample was then centrifuged and an aliquot (100 µL) taken from the supernatant. This aliquot was digested using trypsin (Promega) and analysed using both MALDI and ESI-MS. A trapped carbamate was identified on the N-terminal peptide of the Hb β-chain.

2-2.2.7 Ribulose-1,5-bisphosphate carboxylase oxygenase trapping

Ribulose-1,5-bisphosphate carboxylase oxygenase (RuBisCO) (2 mg, 3.57 nmol) was dissolved in phosphate buffer (2 mL, 50 mM, pH 7.4). $NaHCO_3$ (1.7 mg, 0.02 mmol) was dissolved in phosphate buffer (1 mL, 50 mM, pH 7.4) and added to the protein solution. This solution was added to the pH stat. TEO (280 mg, 1.47 mmol) was added to the reaction mixture in three portion-wise steps in dH_2O (1 mL) while the pH was maintained with automated addition of NaOH (1 M). The reaction mixture was stirred for 1 h then dialysed overnight (1 L dH_2O). The sample was then centrifuged and an aliquot (100 µL) taken from the supernatant. This aliquot was digested using trypsin and analysed using both MALDI and ESI-MS. A trapped carbamate was identified at site K183 but no carbamate was found on the known literature site K201.

2-2.2.8 *Arabidopsis thaliana* leaf lysate trapping

Arabidopsis leaf lysate was prepared as described refer to section 2-3.4. Extracted proteins were re-dissolved in phosphate buffer (2 mL, 50 mM, pH 7.4). $NaHCO_3$ (1.7 mg, 0.02 mmol) was dissolved in phosphate buffer (1 mL, 50 mM, pH 7.4) and added to the protein solution. This solution was added to the pH stat. TEO (280 mg, 1.47 mmol) was added to the reaction mixture in three portion-wise steps in dH_2O (1 mL) while the pH was maintained with automated addition of NaOH (1 M). The reaction mixture was stirred for 1 h then dialysed overnight (1 L dH_2O). The sample was then centrifuged and an aliquot (100 µL) taken from the supernatant. This aliquot was digested using trypsin and analysed using ESI-MS. Several carbamate hits were found using this method and are described in detail in chapter 5.

2-2.2.9 Purified Fructose Bisphosphate Aldolase 1 (FBA1) trapping

FBA1 protein (1 mg, 23.8 nmol) was dissolved in phosphate buffer (2 mL, 50 mM, pH 7.4). $NaHCO_3$ (1.7 mg, 0.02 mmol) was dissolved in phosphate buffer (1 mL, 50 mM, pH 7.4) and added to the protein solution. This solution was added to the pH stat. TEO (280 mg, 1.47 mmol) was added to the reaction mixture in three portion-wise steps in dH_2O (1 mL) while the pH was maintained with automated addition of NaOH (1 M). The reaction mixture was stirred for 1 h then dialysed overnight (1 L dH_2O). The sample was then centrifuged and an aliquot (100 µL) taken from the supernatant. This sample was digested using trypsin and analysed using ESI-MS. A trapped carbamate was identified at site K293 matching the leaf lysate screen results. Details of this are contained in chapter 6.

2-3 Molecular Biology

2-3.1 Synthesis of acetyl-lysine carbamate

Acetyl-lysine (50 mg, 0.25 mmol) and $NaHCO_3$ (50 mg, 0.60 mmol) was dissolved in dH_2O (1 mL). Ethyl chloroformate (27.13 mg, 0.25 mmol) in THF (3 mL) was added with stirring. The reaction was stirred overnight and then reduced under pressure. The precipitate was dissolved in acidified dH_2O (5 mL, pH 2) and the product extracted into diethyl ether (2 × 5 mL) (27.7 mg, 40 % yield). ^{1}H NMR (400 MHz, D_2O) δ/ppm 4.29 (1H, dd, J= 9.0, 5.0 Hz C<u>H</u>NH), 4.07 (2H, q, J= 7.2 Hz C<u>H</u>$_2$CH$_3$), 3.11 (2H, t, J= 6.6 Hz C<u>H</u>$_2$NH), 2.03 (3H, s CH_3CO), 1.90-1.68 (2H, m C<u>H</u>$_2$CH), 1.50 (2H, p, J= 6.8 Hz C<u>H</u>$_2$CH$_2$CH), 1.45-1.32 (2H, m, C<u>H</u>$_2$CH$_2$NH), 1.21 (3H, t, J= 7.1 Hz C<u>H</u>$_3$CH$_2$).

2-3.2 Protein digestion protocol

Dithiothreitol (DTT) (0.77 g, 5 mmol) was added to protein sample (100 µg, 100 µL, 200 mM ammonium bicarbonate buffer pH 8, 10 % MeCN) and incubated at 56 °C for 20 min. The sample was then cooled before iodoacetamide (2.77 mg, 0.015 mmol) was added and the sample incubated at room temperature in the dark for 15 min. Trypsin (Promega, sequencing grade) (2 µg in 10 µL of 2 mM HCl 10 % (v/v) acetonitrile) was activated by heating at 37 °C for 20 min. The activated trypsin was added to the protein sample and incubated overnight at 37 °C.

2-3.3 *Arabidopsis thaliana* plant growth

Arabidopsis seeds were plated onto 0.8 % (w/v) plant agar containing 4.4 g/L Murashige and Skoog salt mixture and incubated at 4 °C for 48 h in the dark. The seeds were then incubated at 22 °C with 12 h of light for one week before planting. *Arabidopsis* plants were transferred to jiffy pellet soil plugs (LBS Horticulture) and grown at 22 °C with 12 h of daylight for 5 weeks before leaf proteins were extracted for experimentation.

2-3.4 *Arabidopsis* protein extraction

Arabidopsis leaves (5 g dry weight) were added to a pestle and mortar and ground in the presence of liquid N_2. Pre-chilled extraction phosphate buffer (100 mM, 15 mL, pH 7.4) was added to the leaves with sand and poly(vinylpolypyrrolidone) (PVPP) and further grinding was carried out. The solution was passed through Miracloth (Millipore) on ice to remove particulates. The filtrate was then centrifuged at 4500 rpm for 10 min at 4 °C. The supernatant from this spin contains the soluble proteins and was used for *Arabidopsis* trapping experiments, with the addition of protease inhibitor cocktail (cocktail for plant cell and tissue extracts, 1% (v/v)). The protein concentration of the plant extract was measured individually for each experiment using a Bradford assay and ~3 mg of protein was used for trapping.

2-3.5 Bradford assay

Samples were made with bovine serum albumin (BSA) standards from 2 mg to 25 µg protein. Bradford reagent was added (1 mL to 20 µL sample) and the samples incubated for 15 min at rt. The absorbance was measured at 595 nm. These measurements were plotted to generate a standard curve. This curve could then be used to measure unknown sample protein concentrations within this range.

2-3.6 Bicinchoninic acid (BCA) assay

Samples were made with bovine serum albumin (BSA) standards from 2 mg to 25 µg protein. The samples (25 µL) were then incubated with a working reagent (Pierce^{TM}) (200 µL) in a microtitre plate for 30 min at 37 °C and then the absorbance was measured at 562 nm. The blank measurement was subtracted from the other readings and the measurements plotted as a standard curve. This curve could then be used to measure unknown sample protein concentrations within this range.

2-3.7 Immunocapture of RuBisCO

Arabidopsis leaf lysate (500 µL) was added to spin column containing RuBisCO antibody and rotated at room temperature for 15 min. The column was centrifuged (30 s, 2000 rpm) and the flow through collected. This RuBisCO depleted sample was then run on a gel for comparison to starting material.

2-3.8 Ammonium sulphate precipitation

Saturated ammonium sulphate was added to *Arabidopsis* leaf lysate (5 mL). At 20 %, 30 % and 40 % (v/v) saturated ammonium sulphate solution the sample was centrifuged (5500 rpm, 10 min) and the supernatant removed to further precipitate. The pellets were dialysed into phosphate buffer (1 L, 50 mM, pH 7.4) and used for trapping experiments.

2-3.9 Protein solubilisation test

Ammonium bicarbonate (ABC) buffer (various concentrations, 100 µL) was added to protein pellets (100 µg) with additives for solubilisation (10 % MeCN or 0.1 % SDS) post trapping reaction. These samples were then incubated at 37 °C (or 95 °C for the heating sample) for 1 h before centrifugation at 13,000 rpm for 5 min. The protein concentration assay was carried out on the supernatant.

2-3.10 Acetone precipitation

Cold (−20 °C) acetone was added to 4× the volume of the protein sample and the reaction cooled to −20 °C for 1 h. This was then centrifuged at 13,000 rpm for 10 min and the pellet washed with cold 80 % (v/v) acetone and re-centrifuged. The residual acetone was allowed to evaporate by incubating uncapped for 30 min at room temperature. The sample proteins were precipitated in a pellet.

2-3.11 Transforming cells

Plasmid DNA (1 µL, 50 ng/µL) was added to competent cells (20 µL). This was mixed by gentle vortex and incubated on ice for 20 mins. This solution was then heat shocked at 42 °C for 45 s and placed back on ice for 2 min. Luria broth (LB) (900 µL) was added and the cells were incubated with shaking at 37 °C for 1 h. The culture was then centrifuged at 6000 rpm for 5 min and the pellet re-suspended in LB (100 µL). This was spread on an LB agar plate containing antibiotics (cell line and plasmid specific) and incubated at 37 °C overnight.

2-3.12 Protein test expression

LB broth (5 mL) containing antibiotics (cell line specific) was inoculated with a plate colony and incubated with shaking at 37 °C overnight. Some of the overnight culture (1 mL) was used to inoculate LB broth (25 mL) containing antibiotics (cell line specific) and incubated with shaking at 37 °C for 3 h until the OD_{600} reached 0.5. A sample was taken of this pre-induction culture (1 mL) and kept for gel comparison. The remaining culture was induced with Isopropyl β-D-1-thiogalactopyranoside (IPTG) (1 mM) and shaken at 37 °C for 3 h. After this time the culture was centrifuged and the pellet re-suspended in lysis buffer (50 mM Tris-HCl pH 7.5, 500 mM NaCl, 30 mM Imidazole, 2 mM β-mercaptoethanol). This was then sonicated at 15 % for 15 s twice and then centrifuged at 5500 rpm for 5 min. Both the supernatant and pellet were run against the pre-induction sample on an SDS-PAGE gel.

2-3.13 SDS-PAGE gel

Protein molecular weights were assessed using sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) gel. Protein samples (10-20 µg) were made up 1:1 in loading buffer (100 mM Tris-HCl pH 6.8, 200 mM dithiothereitol, 4 % (w/v) SDS, 0.2 % (w/v) bromophenol blue, 20 % (v/v) glycerol) and were incubated for 5 min at 95 °C to enable protein denaturation. The samples were run on resolving gels in the range 12 – 15 % (w/v) with 5 % stacking gel. A

protein ladder (Pageruler^{TM} pre-stained) was used to estimate protein size. The gels were run in running buffer (25 mM Tris-HCl pH 7.5, 192 mM glycine, 0.1 % (w/v) SDS) at 180 V for 1 h. The gel was incubated with Coomassie Brilliant Blue G (20 mL, 3 mM Coomassie Brilliant Blue G, 12 M methanol and 2 M glacial acetic acid) with rocking overnight.

2-3.14 Truncation of FBA1 DNA

Forward primer sequence (ggccatatgGCgagcgcgtacgcggacg) from MWG Eurofins was used in conjunction with reverse T7 primer. These were used in the order of components described in the PCR section 2-3.15. The DNA from this PCR was purified from an agarose gel and ligated into pJET1.2 vector as described 2-3.18.

2-3.15 Polymerase chain reaction (PCR)

The components and conditions for PCR to amplify purified DNA are listed in Table 1-1 reaction components and conditions for the amplification of purified DNA.The components were added in the described order. Colony PCR required additional $MgCl_2$ (1 µL) and the plasmid DNA was substituted for a plate colony. The conditions were the same except the extension time was increased to 1 min.

Table 1-1 reaction components and conditions for the amplification of purified DNA.

PCR reaction components		**PCR reaction**		
Component	µL	Step	Temp. °C	Time
H_2O	14.1	1	98	2 min
DMSO	1	2 (×30)	98	20s
GC Buffer	5		59	20s
2 mM dNTPs	2.5		72	30s
Plasmid DNA	0.2	3	72	2 min
Fw primer (25 pmol/µL)	1	4	4	Hold
Rv primer (25 pmol/µL)	1			
Enzyme	0.2			

2-3.16 Agarose gel

Agarose (1.5 g) was dissolved in TBE buffer (100 mL) and microwaved for 1 min. Ethidium bromide (0.1 mg) was added before the gel was poured and allowed to set. DNA samples were run (25 µL) on the gel at 120 V for 1 h with TBE buffer, resulting bands were visualised under UV transillumination.

2-3.17 DNA purification from agarose gel

Prep-a-gene buffers used for DNA purification were purchased from Bio-rad. Agarose gel bands were cut out and melted in binding buffer (1 mL) at 68 °C. Silica (166 mg/mL, 12 µL) was added to the melted gel and the mixture rotated at room temperature for 30 min. This was then centrifuged at 13,000 rpm for 60 s and the supernatant removed. The silica was rinsed with binding buffer (125 µL) and centrifuged. The silica was then washed twice with wash buffer (750 µL) with centrifugation and supernatant removal in between wash steps. The remaining ethanol was evaporated by heating the open tube at 68 °C for 5 min. The silica was then incubated with dH_2O (12 µL) at 37 °C for 10 min and the DNA eluted by centrifugation for 60 s and removal of the supernatant.

2-3.18 Ligation of DNA into a vector

During this work DNA was ligated into either a pJET1.2 or pET-14b vector depending on the nature of the DNA work. For blunt end ligation into a pJET1.2 vector DNA produced from the PCR reaction (3.7 µL) was incubated with ligation buffer (50 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 1 mM ATP, 10 mM DTT, 5 µL), pJET1.2 vector (0.3 µL) and DNA ligase (1 µL) at room temperature for 1 h. This ligation mixture (2 µL) was added to DH5α competent cells (20 µL) and transformed as described 2-3.11. For a pET14b vector the vector (1 µL) and insert (3 µL) were incubated with dH_2O (3 µL), buffer (1 µL), 10 mM ATP (1 µL) and DNA ligase (1 µL) at room temperature for 1 h. This ligation mixture (2 µL) was added to DH5α competent cells (20 µL) and transformed as described 2-3.11.

2-3.19 Mini-prep of overnight cultures

Mini preps were carried out using a miniprep spin kit (QIAGEN). Overnight LB culture (5 mL) was centrifuged at high speed for 5 min and the supernatant removed. The pellet was re-suspended in resuspension buffer (250 µL), and transferred to a microcentrifuge tube. Lysis solution (250 µL) was added and the tube inverted 3-4 times, before being incubated for 2 min. Neutralisation buffer was

added (350 µL) and the tube inverted 4-6 times. The mixture was then centrifuged at 13000 rpm for 5 min and the supernatant transferred to a spin column. This was centrifuged for 1 min and the flow through discarded. Wash solution (500 µL) was added, the column centrifuged, the flow through discarded, and then repeated with an empty column to remove residual solution. The column was transferred to a fresh collection tube and incubated with dH_2O (50 µL), at room temperature for 20 min and centrifuged for 2 min to collect the DNA.

2-3.20 Digestion of vector and inserts

A digestion mixture was made using DNA (25 µL), CutSmart buffer (New England BioLabs) (3 µL) and digestion enzymes Nde1 (1 µL) and BamH1 (1 µL). The mixture was incubated at 37 °C for 1 h. SAP (shrimp alkaline phosphatase) (1 µL) was added to the vector digest and incubated for 30 min at 37 °C and then 10 min 75 °C. These incubations are then run on an agarose gel (2-3.16) to determine the success of the digest and then purified as described previously (2-3.17).

2-3.21 Large scale protein growth

An overnight culture (250 mL) was inoculated with a plate colony and the culture incubated with shaking at 37 °C overnight. Some of this overnight culture (40 mL) was then used to inoculate large scale flasks (1 L LB broth) containing antibiotics (cell line specific). The cultures were incubated with shaking at 37 °C for 2 h until an OD_{600} of 0.5. The cultures were then cooled to 17 °C and induced with IPTG (final concentration 1 mM) and grown for 24 h at 17 °C with shaking. The culture was then centrifuged at 4000 rpm for 20 min and the pellet stored at −80 °C until purification. The protein of interest was purified as described in 2-3.22.

2-3.22 Protein purification

A bacterial cell pellet (20 mL) was re-suspended in binding buffer (35 mL, 50 mM Tris pH 7.5, 100 mM NaCl, 10 mM Imidazole) and sonicated at 40 % for 1 min three

times. This solution was then centrifuged for 1 h at 21000 rpm and the supernatant incubated on Ni^{2+} resin with mixing for 1 h. This resin was then added to a column and allowed to settle with the flow through collected. The column was then rinsed with a column volume of binding buffer and collected. The column was then washed with 3 column volumes of wash buffer (35 mL, 50 mM Tris pH 7.5, 100 mM NaCl, 10 mM Imidazole), with fractions collected of each column volume. The protein was then eluted with increasing amounts of imidazole in binding buffer (range 50 mM to 250 mM). Elution buffer was passed through the column until no more protein was detected at 280 nm in the eluting liquid. All the column fractions were then run on an SDS-PAGE gel (2-3.13).

2-3.23 One-step site directed mutagenic PCR

PCR was carried out using the same conditions as 2-3.15 but also containing the template DNA and with 20 cycles of step 2 table 2-1 melting, annealing, extending to give additional time to replicate the whole vector. After the PCR reaction Dpn1 (1 µL) was added to the PCR mixture (20 µL) and incubated at 37 °C for 4 h. This mixture was then used to transform DH5α cells 2-3.11.

2-3.24 Refolding from inclusion bodies

Cells were lysed in binding buffer (35 mL, 50 mM Tris pH 7.5, 100 mM NaCl, 10 mM Imidazole) and centrifuged (21,000 rpm, 40 min). The supernatant is discarded and the pellet resuspended in lysis buffer (50 mM Tris-HCl pH 7.5, 500 mM NaCl, 30 mM Imidazole, 2 mM β-mercaptoethanol) and centrifuged (21000 rpm, 10 min) and the supernatant discarded (repeat three times). The pellet was then washed with binding buffer (20 mL) then binding buffer containing 8 M urea was added (10 mL) and the sample rotated for 30 min. The sample in urea was diluted 1 in 50 into refolding buffer (50 mM Tris-HCl pH 8.5, 240 mM NaCl, 10 mM KCl, 1 mM EDTA, 0.5 % triton X-100, 1 mM DTT) with shaking and the activity measured to monitor refolding.

2-3.25 FBA1 cleavage assay

Purified FBA1 protein (10 µg) was added to a mixture of β-NADH (0.2 mM), EDTA (10 mM) and coupling enzyme α-GDH/TPI (10 units) in TBE buffer (40 mM, pH 7.5). Substrate fructose-1,6-bisphosphate (2 mM) was added and the reaction mixed for 30 s before absorbance read at 340 nm every 5 min over 20 min. For CO_2 minus experiments: 96-well plate incubated under Ar atmosphere for 30 min prior to F-1,6-BP substrate addition. The enzyme reaction was then stopped with the addition of 5 % TFA.

2-4 Physical Chemistry

2-4.1 Concentrating a digest sample on a ziptip

Samples were adjusted to 0.1 % (v/v) TFA. The ziptip was washed with 100 % acetonitrile (2 × 10 µL) and then equilibrated with 0.1 % TFA (3 × 10 µL). The sample was then bound to the tip (30 µL) and expirated 7 – 10 times. The tip was then washed with 0.1 % (v/v) TFA and the peptides eluted with 0.1 %/TFA 50 % acetonitrile (v/v) (10 µL).

2-4.2 Fractionating on a StageTip

Adapted from literature (135). The stage tip was formed of a C^{18} layer above a strong cation exchange membrane. The membranes were washed with methanol (20 µL) and then equilibrated with formic acid. The sample (20 µL made up with 0.1 % TFA) was added to the tip and fractions collected with increasing ammonium acetate concentrations ranging from 50 – 500 mM (10 µL elutions).

2-4.3 Accurate mass

Protein samples were dialysed overnight against ultra-pure water (1 L) and then resuspended 50 µL protein (100 µg) with 50 µL acetonitrile and 1 % formic acid. Accurate mass was obtained on an LTQ FT machine.

2-4.4 MALDI-MS

Peptide samples (1 µL) were applied to a MALDI plate in the ratio 1:1 with α-Cyano-4-hydroxycinnamic acid (α-CHCA, saturated solution in 50:50 water/acetonitrile with 0.1 % TFA) matrix (1 µL). The spots were allowed to crystallise through evaporation at room temperature before placing the plate into the MALDI ion source. Data was

collected using 100 pulse shots with variable laser intensity depending on the sample concentration.

2-4.5 ESI-MS method

Samples were prepared in MS compatible ABC buffer. These samples were run with a 2 h gradient of acetonitrile from 2-80 % containing 0.1 % formic acid.

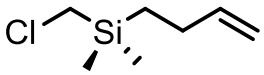
2-4.6 Data analysis

The ESI data was analysed using the GPM database X!Tandem (Tandem) (136) or MaxQuant (137). Modifications for the transfer of an ethyl group (MW 28.0313) and a trapped carbamate group (MW 72.0211) were searched from among a protein database created from *Arabidopsis* peptides that were found in the results.

2-5 Synthetic chemistry

2-5.1 Synthesis of (But-3-enyl)dimethylsilane halogens

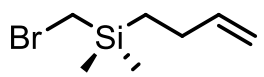
2-5.1.1 Chloromethyl (But-3-enyl)dimethylsilane



4

Following a literature procedure (138), 4-bromo-1-butene (7.98 g, 59.1 mmol) in anhydrous diethyl ether (5 mL) was added dropwise to magnesium (1.56 g, 65 mmol) and dibromoethane (0.1 mL) in anhydrous diethyl ether (10 mL). The mixture was stirred at 23 °C for 3 h. Chloro(chloromethyl)dimethylsilane (3.58 g, 25 mmol) in anhydrous diethyl ether (5 mL) was then added and the mixture was refluxed at 45 °C for 39 h. The reaction was quenched by slow addition of dH_2O (5 mL) and solid residues were removed by vacuum filtration. The dH_2O layer was separated and the ethereal layer was washed with saturated potassium carbonate solution (2 × 10 mL), dried ($MgSO_4$) and concentrated under reduced pressure. Purification by column chromatography (silica gel, hexanes) yielded the chloro alkene **4** (X=Cl) (1.85 g, 55%); ^{1}H NMR (400 MHz, $CDCl_3$) δ/ppm 5.86 (1H, ddt, J=16.6, 10.1, 6.3 Hz, CH=), 5.04-4.98 (1H, m, C<u>H</u>H=), 4.93-4.90 (1H, m, CH<u>H</u>=), 2.79 (2H, s, CH_2Cl), 2.14-2.07 (2H, m, CH_2CH), 0.78-0.74 (2H, m CH_2Si), 0.12 (6H, s, CH_3Si). ^{13}C NMR (100 MHz, $CDCl_3$) δ/ppm 140.8, 113.2, 30.2, 27.6, 12.8 −4.7.

2-5.1.2 Bromomethyl (But-3-enyl)dimethylsilane



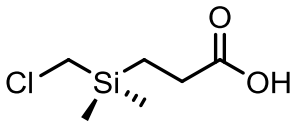
4

Adapting a literature procedure (138), 4-bromo-1-butene (27.9 g, 206.9 mmol), in anhydrous diethyl ether (30 mL) was added dropwise to magnesium (4.3 g, 180 mmol) and dibromoethane (0.1 mL) in anhydrous diethyl ether (130 mL). The mixture was stirred at 23 °C for 3 h. Chloro(bromomethyl)dimethylsilane (22 g, 117.3 mmol) in anhydrous diethyl ether (30 mL) was then added and the mixture was

refluxed at 45 °C for 24 h. The reaction was quenched by slow addition of saturated ammonium chloride solution (80 mL). The product was extracted into diethyl ether (2 × 30 mL), and the ethereal layer was washed with saturated potassium carbonate solution (2 × 20 mL), dried ($MgSO_4$) and concentrated under reduced pressure. Purification by column chromatography (silica gel, hexanes) yielded the bromo alkene **4**(X=Br) (11.7 g, 58%); as a colourless liquid. ^{1}H NMR (400 MHz, $CDCl_3$) δ/ppm 5.87 (1H, ddt J= 16.6, 10.1, 6.3 Hz, CH=), 5.04-4.99 (1H, m, C<u>H</u>H=), 4.94-4.90 (1H, m, CH<u>H</u>=), 2.48 (2H, s, CH_2Br), 2.12-2.07 (2H, m, CH_2CH), 0.80-0.75 (2H, m, CH_2Si), 0.14 (6H, s, CH_3Si). ^{13}C NMR (100 MHz, $CDCl_3$) δ/ppm 140.9, 113.2, 27.7, 17.0, 13.2, −4.0. EI-GC MS: 150.9 (25, $[M - C_4H_7]^+$), 122.9 (26, $[M - C_6H_{12}]^+$), 113.1 (100, $[M - CH_2Br]^+$), 85.0 (24), 59.0 (22). IR 1640.

2-5.2 Synthesis of silyl propanoic acid halogens

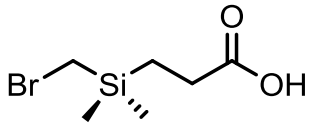
2-5.2.1 3-[(chloromethyl)dimethylsilyl]propanoic acid



5

Adapting a literature procedure (138), sodium periodate (3.8 g, 1.78 mmol) was added to chloro alkene **1** (5.86 g, 36 mmol) in a mixture of chloroform (25 mL), acetonitrile (25 mL) and dH_2O (40 mL) and the mixture was heated to 50 °C for 10 min. Ruthenium III chloride hydrate (19.3 mg, 0.09 mmol) was added and the mixture was stirred for 1 h. Further sodium periodate (9.5 g, 4.44 mmol) was added portion-wise over 3 h. The mixture was diluted with dichloromethane (10 mL) and dH_2O (20 mL) and the phases were separated. The dH_2O layer was extracted with CH_2Cl_2 (3 × 10 mL) and the organic layers were combined and dried ($MgSO_4$). The extracts were concentrated under pressure and purified by column chromatography (silica, 1:2 hexane:ethyl acetate) to yield the chloro acid **5**(X=Cl) liquid (2.42 g, 37%); ^{1}H NMR (400 MHz, $CDCl_3$) δ/ppm 2.80 (2H, s, CH_2Cl), 2.43-2.39 (2H, m, CH_2C=O), 1.02-0.98 (2H, m, CH_2Si), 0.14 (6H, s, CH_3Si). ^{13}C NMR (100 MHz, $CDCl_3$) δ/ppm 180.9, 29.7, 28.3, 8.5, −4.9.

2-5.2.2 3-[(bromomethyl)dimethyl)silyl]propanoic acid

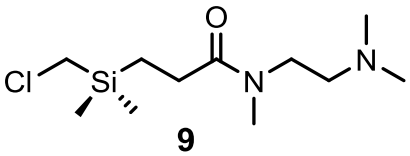


5

Adapting a literature procedure (138), sodium periodate (1.9 g, 0.89 mmol) was added to bromo alkene **4** (0.3 g, 1.84 mmol) in a solution of chloroform (5 mL), acetonitrile (5 mL) and dH_2O (8 mL) and the mixture was heated to 50 °C for 10 min. Ruthenium III chloride hydrate (10 mg, 0.05 mmol) was added and the mixture was stirred for 3 h. Further sodium periodate (4.5 g, 20.9 mmol) was added stepwise over 33 h. The mixture was diluted with dichloromethane (10 mL) and dH_2O (20 mL) and the phases separated. The dH_2O layer was extracted with CH_2Cl_2 (3 × 10 mL) and the organic layers were combined and dried ($MgSO_4$). The extracts were concentrated under pressure to yield bromo acid **5**(X=Br) (0.31 g, 98%); yellow liquid. ^{1}H NMR (400 MHz, $CDCl_3$) δ/ppm 2.47 (2H, s, CH_2Br), 2.42-2.37 (2H, m, $CH_2C=O$), 1.02-0.98 (2H, m, CH_2Si), 0.15 (6H, s, CH_3Si). ^{13}C NMR (100 MHz, $CDCl_3$) δ/ppm 181.1, 28.4, 16.2, 9.0, −4.2. EI-GC MS: 129.07 (21, $[M - BrOH]^+$), 101.08 (34, $[M – C_2O_2HBr]^+$), 88.05 (54), 58.00 (100), 43.05 (52). IR 1705.

2-5.3 Synthesis of amide halogens

2-5.3.1 3-[(chloromethyl)(dimethyl)silyl] [(methylamino)(dimethylamino)ethyl] Propamide

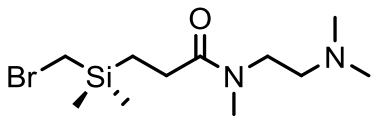


9

Adapting a literature procedure (139), thionyl chloride (0.70 g, 5.51 mmol) was added to chloro acid **5** (0.31 g, 1.74 mmol) and the mixture was stirred at 60 °C for 2 h. The excess thionyl chloride was removed under reduced pressure,

trimethylethylenediamine (0.08 g, 0.77 mmol) was added and the mixture was stirred for 30 min. The reaction mixture was diluted with diethyl ether (10 mL) and washed with saturated potassium carbonate solution (8 mL), the organic layer was dried ($MgSO_4$) and concentrated under vacuum to yield chloro amide **9** solid (0.16 g, 36%); ^{1}H NMR (400 MHz, $CDCl_3$) δ/ppm (trans conformer) 3.48 (2H, t, J= 7.0 Hz, CH_2NH_3), 3.01 (3H, s, NH_3), 2.81 (2H, s, CH_2Cl), 2.54-2.42 (2H, m, $CH_2N(CH_3)_2$), 2.37-2.33 (2H, m, $CH_2C=O$), 2.27 (6H, s, $N(CH_3)_2$), 1.00-0.95 (2H, m, CH_2Si), 0.13 (6H, s, CH_3Si). ^{1}H NMR (400 MHz, $CDCl_3$) δ/ppm (cis conformer) 3.37 (2H, t, J= 7.0 Hz, CH_2NH_3), 2.93 (3H, s, NH_3), 2.81 (2H, s, CH_2Cl), 2.54-2.42 (2H, m, $CH_2N(CH_3)_2$), 2.37-2.33 (2H, m, $CH_2C=O$), 2.27 (6H, s, $N(CH_3)_2$), 1.00-0.95 (2H, m, CH_2Si), 0.13 (6H, s, CH_3Si). ^{13}C NMR (100 MHz, $CDCl_3$) δ/ppm 173.5, 48.3, 45.6, 35.8, 34.1, 30.4, 27.8, 27.1, 8.8, −4.5. IR 1636.

2-5.3.2 [(bromomethyl)(dimethyl)silyl] [(methylamino)3-(dimethylamino)ethyl] Propamide



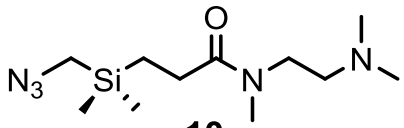
9

Adapting a literature procedure (139), thionyl chloride (0.33 g, 2.76 mmol) was added to bromo acid **5** (0.33 g, 1.47 mmol) and the mixture was stirred at room tempterature for 10 min. The excess thionyl chloride was removed under reduced pressure before the addition of trimethylethylenediamine (0.14 g, 1.38 mmol) with stirring for 30 min. The reaction mixture was diluted with diethyl ether (10 mL) and washed with saturated potassium carbonate solution (8 mL), the organic layer was dried ($MgSO_4$) and concentrated under vacuum to yield bromo amide **9** (0.44 g, 97%); ^{1}H NMR (400 MHz, $CDCl_3$) δ/ppm (trans conformer) 3.46 (2H, t, J= 7.0 Hz, CH_2NCH_3), 3.00 (3H, s, NCH_3), 2.49 (2H, s, CH_2Br), 2.45-2.41 (2H, m, CH_2N), 2.37-2.33 (2H, m, $CH_2C=O$), 2.26 (6H, s, $N(CH_3)_2$). 1.02-0.93 (2H, m, CH_2Si), 0.14 (6H, s, $Si(CH_3)_2$). ^{1}H NMR (400 MHz, $CDCl_3$) δ/ppm (cis conformer) 3.35 (2H, t, J= 7.0 Hz, CH_2NCH_3), 2.92 (3H, s, NCH_3), 2.49 (2H, s, CH_2Br), 2.45-2.41 (2H, m, CH_2N), 2.37-2.33 (2H, m, $CH_2C=O$), 2.26 (6H, s, $N(CH_3)_2$). 1.02-0.93 (2H, m, CH_2Si), 0.14

(6H, s, $Si(CH_3)_2$). ^{13}C NMR (100 MHz, $CDCl_3$) δ/ppm 173.4, 48.2, 45.8, 45.5, 35.7, 34.0, 27.7, 17.0, 9.2, −4.0. IR 1670.

2-5.4 Synthesis of amide azide

[(azidomethyl)(dimethyl)silyl] [(methylamino)3-(dimethylamino)ethyl] Propamide

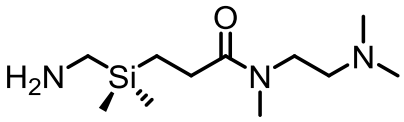


10

Adapted from literature (140) bromo amide **9** (0.25 g, 0.81 mmol) was dissolved in acetone (2 mL) and a solution of sodium azide (0.26 g, 4.03 mmol) in dH_2O (0.5 mL) was added. The solution was then stirred for 48 h at room temperature. The acetone was removed under pressure and diluted with potassium carbonate solution (4 mL). The reaction was then washed with dichloromethane (2 × 10 mL). The dichloromethane layers were combined, dried ($MgSO_4$) and reduced under pressure to yield azide amide **10** (92%)(0.20 g, 0.74 mmol); ^1H NMR (400 MHz, $CDCl_3$) δ/ppm ppm (trans conformer) 3.48 (2H, t, J= 7.0 Hz, CH_2NCH_3), 3.01 (3H, s, NCH_3), 2.49 (2H, s, CH_2N_3), 2.45-2.41 (2H, m, CH_2N), 2.37-2.33 (2H, m, $CH_2C=O$), 2.26 (6H, s, $N(CH_3)_2$). 1.02-0.93 (2H, m, CH_2Si), 0.14 (6H, s, $Si(CH_3)_2$). ^1H NMR (400 MHz, $CDCl_3$) δ/ppm (cis conformer) 3.36 (2H, t, J= 7.0 Hz, CH_2NCH_3), 2.93 (3H, s, NCH_3), 2.49 (2H, s, CH_2N_3), 2.45-2.41 (2H, m, CH_2N), 2.37-2.33 (2H, m, $CH_2C=O$), 2.26 (6H, s, $N(CH_3)_2$). 1.02-0.93 (2H, m, CH_2Si), 0.14 (6H, s, $Si(CH_3)_2$). ^{13}C NMR (100 MHz, $CDCl_3$) δ/ppm 173.4, 48.4, 46.0, 45.7, 41.4, 36.0, 34.3, 27.8, 9.2, −4.0. IR 2091.

2-5.5 Synthesis of amide amine

[(aminemethyl)(dimethyl)silyl] [(methylamino)3-(dimethylamino)ethyl] Propamide

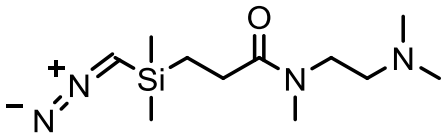


13

To **10** (0.09 g, 0.34 mmol) was added triphenylphosphine (0.09 g, 0.34 mmol) in THF (1 mL) with H_2O (5 µL). This reaction was stirred at room temperature for 48 h. This mixture was washed with basic H_2O (containing potassium carbonate, pH 10) and then extracted with DCM (3 × 5 mL). This solution was reduced under pressure to yield amine amide (**13**) (0.11 g, 0.46 mmol). ^{1}H NMR (400 MHz, $CDCl_3$) δ/ppm ppm (trans conformer) 3.54-3.49 (2H, m, CH_2NCH_3), 3.02 (3H, s, NCH_3), 2.53-2.49 (2H, m, CH_2N), 2.45-2.42 (2H, m, $CH_2C{=}O$), 2.27-2.26 (8H, m, NH_2 and $N(CH_3)_2$). 0.92-0.85 (2H, m, CH_2Si), 0.13 (6H, s, $Si(CH_3)_2$). ^{1}H NMR (400 MHz, $CDCl_3$) δ/ppm (cis conformer) 3.4-3.35 (2H, m, CH_2NCH_3), 2.94 (3H, s, NCH_3), 2.53-2.49 (2H, m, CH_2N), 2.45-2.42 (2H, m, $CH_2C{=}O$), 2.27-2.26 (8H, m, NH_2 and $N(CH_3)_2$). 0.92-0.85 (2H, m, CH_2Si), 0.13 (6H, s, $Si(CH_3)_2$).

2-5.6 Attempted synthesis of amide diazo compound

[(diazomethyl)(dimethyl)silyl] [(methylamino)3-(dimethylamino)ethyl] Propamide



1

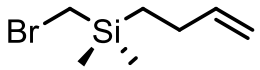
Adapted from a literature procedure used to make TMS-DAM (141). Dry diethyl ether (2 mL) and dibromomethane (0.01 mL, 0.11 mmol) were added to magnesium turnings (65 mg, 2.7 mmol) in a dry two-headed flask. The mixture was stirred at room temperature for 15 min. In a separate flask dry diethyl ether (2 mL) was added to **9** (52 mg, 0.2 mmol) and transferred to the reaction mixture. The mixture was stirred at 40 °C for 3 h. The mixture was then added to dry diethyl ether (2 mL) and

diphenylphosphoryl azide (130 mg, 0.47 mmol) on ice. The reaction was quenched with dH_2O (10 mL) and extracted with diethyl ether (2 × 10 mL). The ethereal layer was concentrated under pressure. The reaction was not successful in producing diazo product due to lack of formation of the Grignard.

Adapted from literature procedure (142). Dissolve **10** (0.03 g, 0.12 mmol) in THF (2 mL) and 100 μL dH_2O. Then add *N*-succinimidyl 3-(diphenylphosphino) propionate (0.03 g, 0.09 mmol), stir for 8 h on ice, leave to warm to room temperature overnight. Mixture diluted with sodium chloride and extracted with dichloromethane. The reaction progressed as far as phosphine attachment but the separation to produce the diazo was not achieved.

2-5.7 Conversion of Cl alkene to Br alkene (Finklestein reaction)

Bromomethyl (But-3-enyl)dimethylsilane



4

To **4**(**X**=Cl) (0.57 g, 3.5 mmol) was added lithium bromide (4.4 g, 51.2 mmol) in dry acetone (2 mL) and heated to 40 °C for 40 h. The acetone was removed under reduced pressure and the mixture was extracted with diethyl ether (2 × 5 mL). The extracts were combined and concentrated under reduced pressure to yield bromoalkene **4** (0.27 g, 1.31 mmol); ^{1}H NMR (400 MHz, $CDCl_3$) δ/ppm 5.90 (1H, ddt, J= 16.4, 10.1, 6.3 Hz, CH=), 5.04-4.99 (1H, m, CH_2=), 4.94-4.90 (1H, m, CH_2=), 2.48 (2H, s, CH_2Br), 2.14-2.07 (2H, m, CH_2CH), 0.80-0.75 (2H, m, CH_2Si), 0.14 (6H, s, CH_3Si). ^{13}C NMR (100 MHz, $CDCl_3$) δ/ppm 140.9, 113.2, 27.7, 17.0, 13.2, −4.0.

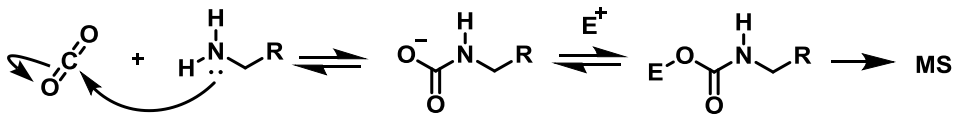
Chapter 3: Design and partial synthesis of a Trialkylsilyl-diazomethane Derivative for Carbamate trapping

3-1 Overview

Protein carbamates are unstable, and do not withstand standard methods for PTM analysis such as MS (85). Previous analyses of protein carbamates were therefore carried out using artificial conditions (106,110,115).

X-ray crystallography has been used to confirm the site of carbamylation on Ribose-1,5-bisphosphate carboxylase oxygenase (RuBisCO) (100). Samples used to form crystals were never tested for their enzymatic activity and the concentrations of CO_2 used far exceeded physiological conditions. NMR spectroscopy has been used to confirm carbamate formation on haemoglobin (Hb) and RuBisCO (143,144). The main issue with this technique is that it is difficult to confirm the site of carbamate formation and would be unusable within a protein mixture due to the complexity of the sample. The carbamate site on Hb was later inferred by modification with cyanate (145).

This thesis describes a strategy to trap carbamates under cellular conditions thereby making them amenable to downstream analysis; this method relies on the nucleophilic attack of carbamates upon electrophiles to introduce stability (Scheme 3-1).



Scheme 3-1 Formation and trapping of a carbamate using an electrophilic reagent.

To achieve this goal two approaches are described, the first is covered in this chapter; the design and partial synthesis of a trimethylsilyl-diazomethane (TMS-DAM) derivative. This work was not completed due to the success of the second approach using commercially available Meerwein's reagents trimethyloxonium tetrafluoroborate (TMO) and triethyloxonium tetrafluoroborate (TEO) salts. The development of that methodology process is detailed in chapter 4.

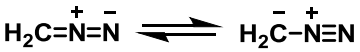
3-2 Background to the trapping hypothesis

Diazomethane and TMS-DAM are established reagents for the preparation of alkyl esters (146). The following section reviews the underlying mechanistic chemistry behind the use of these systems as reagents for carboxylic acid esterification. This literature work forms the basis for the approach and molecule design discussed in this chapter.

Trapping of carbamates is required for their analysis because the labile nature of the modification renders it unsuitable for normal PTM analysis (Scheme 3-1). Often PTM sites can be confirmed by mutagenesis, the site of carbamate formation on Hb was confirmed by modifying the α-amine groups of the *N*-terminal residues using cyanate (147). This is not a technique that could be applied to the majority of suspected carbamylated lysines as they are often located close to or within active sites. In the case of RuBisCO it was known that the site could not be confirmed by mutagenesis because it would render the enzyme inactive (102).

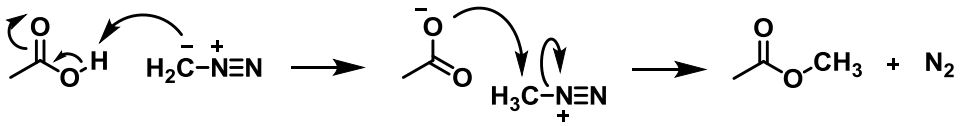
3-2.1 Diazomethane

Diazomethane (DAM) has previously been used to efficiently convert organic acids into methyl esters, where yields are typically 95-100% (148). The diazo group can be represented by two resonance structures (149), where the C atom is basic, making it susceptible to protonation by carboxylic acids (Scheme 3-2).



Scheme 3-2 Resonance structures of diazomethane.

Once protonated, the methyl group becomes electrophilic, allowing irreversible attack by the carboxylate group with the release of nitrogen gas. The process usually proceeds with few side reactions (Scheme 3-3).



Scheme 3-3 Mechanism of esterification of a carboxylic acid by diazomethane.

DAM has previously been used to investigate CO_2 binding to RuBisCO. DAM was added to a RuBisCO complex bound to $^{14}CO_2$, the protein was then denatured using SDS and seen to still contain the $^{14}CO_2$ showing that it had been 'trapped' by the DAM (83).

However, DAM is a highly toxic, explosive gas which has now been given a high carcinogenic rating (150-154), making it a difficult substance to work with safely and only appropriate to be used on small scales. These properties made it necessary to develop a safer alternative.

The reactions of amines with carbon dioxide to form carbamates have been explored outside cellular systems within the context of carbon dioxide scrubbing (155). These reactions differ from the *in vivo* protein systems to be explored because the levels of CO_2 are far higher than physiological, but they provide unequivocal evidence for the reactions between amines and carbon dioxide to form carbamates.

3-2.2 Trimethylsilyl-diazomethane

Trimethylsilyl-diazomethane (TMS-DAM) is a safer working alternative to DAM (156) (Figure 3-1). The addition of the TMS group provides stability to the diazo functionality and the molecule can now be purchased in hexane or diethyl ether as a stable liquid (156).

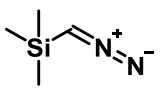
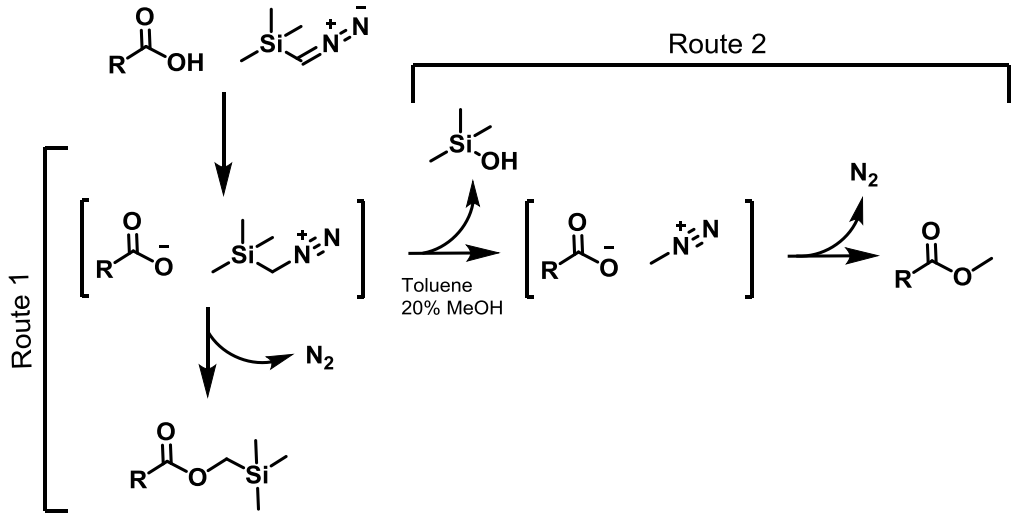


Figure 3-1 Structure of TMS-DAM.

TMS-DAM is an electrophilic methylating agent and has been used previously in organic systems for the *O*-methylation of alcohols in dichloromethane (121) and for the transfer of the CH_2-TMS group to carboxylic acids in benzene (150). This reagent has been compared to other methods of esterification of acids and though it was found to require milder conditions it also yielded TMS ester side-products (Scheme 3-4, route 1) (157). It was later found that if the reaction is carried out with the addition of methanol (20 % methanol in toluene), only the CH_3 ester is formed (Scheme 3-4, route 2) (141,155). This work thus provided an efficient method for methylation of carboxylic acids; however, this work lacked a deeper mechanistic explanation.

In 2007, Kühnel *et. al.* (150) used isotopic labelling to demonstrate that using methanol as a co-solvent causes the removal of the TMS group, to afford DAM *in situ*, which then transfers a methyl group to the carboxylate to afford methyl esters (150) rather than CH_2-TMS esters (Scheme 3-4, route 2).



Scheme 3-4 *In situ* formation of DAM from TMS-DAM followed by transfer of a methyl group to RCO_2^-.

Work using TMS-DAM to trap carbamates has been carried out previously, but only within a mostly organic (benzene/methanol (4:1)) solvent (122). This work involved the bubbling of carbon dioxide into solutions containing amines followed by the

addition of TMS-DAM to produce the corresponding methylcarbamates. These reactions produced good yields, but the organic solvents cannot be used in this work, mainly because any reaction under non-physiological conditions could create artificial results due to the role that water plays in the dynamics of proteins (158).

Carbamates only form on neutral amines. At physiological pH 7.4 most amines are positively charged, therefore carbamates are only expected to form under privileged conditions (i.e. at specific active sites) that have increased concentrations of neutral amine.

These precedents for the application of TMS-DAM towards the trapping of carbamates form the basis for the development of the work described in this chapter. The hypothesis of how to adapt TMS-DAM into a water-soluble reagent is put forward in the design section below (3-3). However this trapping method was developed in parallel with a method utilising Meerwein's reagents TMO and TEO. The use of TEO demonstrated successful results so that method was developed and this synthesis was never fully completed.

3-3 Design of a water-soluble TMS-DAM derivative

Based on the previous work using TMS-DAM for esterification of carboxylic acid and carbamate groups, this molecule was selected as the starting point for the synthetic design. The key feature needed for the molecule to trap a carbamate under physiological conditions was to be water soluble. On this basis, a water-soluble group was introduced into TMS-DAM at position R (Figure 3-2).

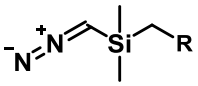
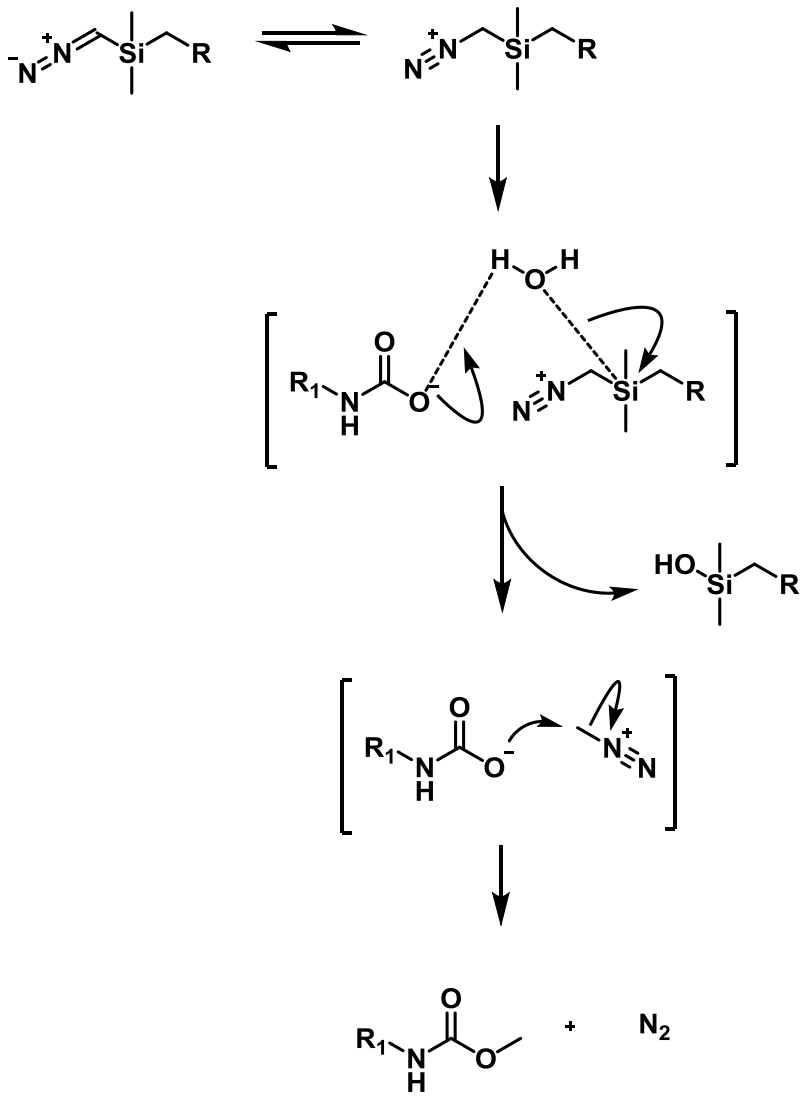


Figure 3-2 Silyl diazo molecule showing the site of introduction of a water-soluble group (R) into TMS-DAM.

The work of Kühnel *et. al.* (150) discussed in 3-2.2 illustrated the requirement for methanol to remove the TMS group. The hypothesis in this work was that within a cellular system water would affect the desilylation process in place of methanol. Scheme 3-5 shows the predicted reaction mechanism of desilylation and carbamate trapping.



Scheme 3-5 Hypothesised reaction mechanism of trapping a carbamate by a water soluble diazo trapping reagent.

The trapping molecule needs to be soluble at neutral pH; the normal method of introducing solubility to a molecule is through the addition of a charged group. Use of an anionic group is not possible because it will provide a nucleophile, which could easily react with the diazo group to become methylated. Therefore a cationic group was necessary; amines are cationically charged at neutral pH and have been used

to increase the solubility of other molecules, such as pro-dugs, within cellular systems (159).

In order to prevent intramolecular reaction between the diazo and the amine functionalities, an amide bond will be used to introduce rigidity to the molecule. Trimethylethylenediamine was chosen because it contains only one site for amide formation and, once attached, will not contain any primary or secondary amine groups, which would also interfere with the trapping reagent.

In order to form the amide bond, a carboxylic acid will need to be introduced into the silyl backbone. This will be achieved via oxidation of an alkene group and is discussed in further detail later in this chapter (3-5.2).

Based on these chemical assumptions, the water-soluble carbamate trapping agent design is summarised below (**1**) (Figure 3-3).

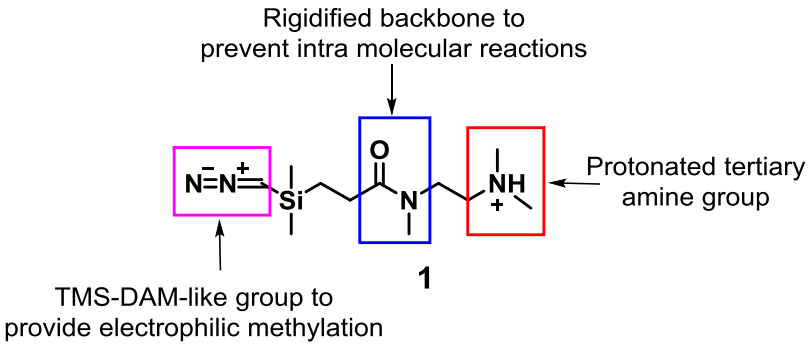
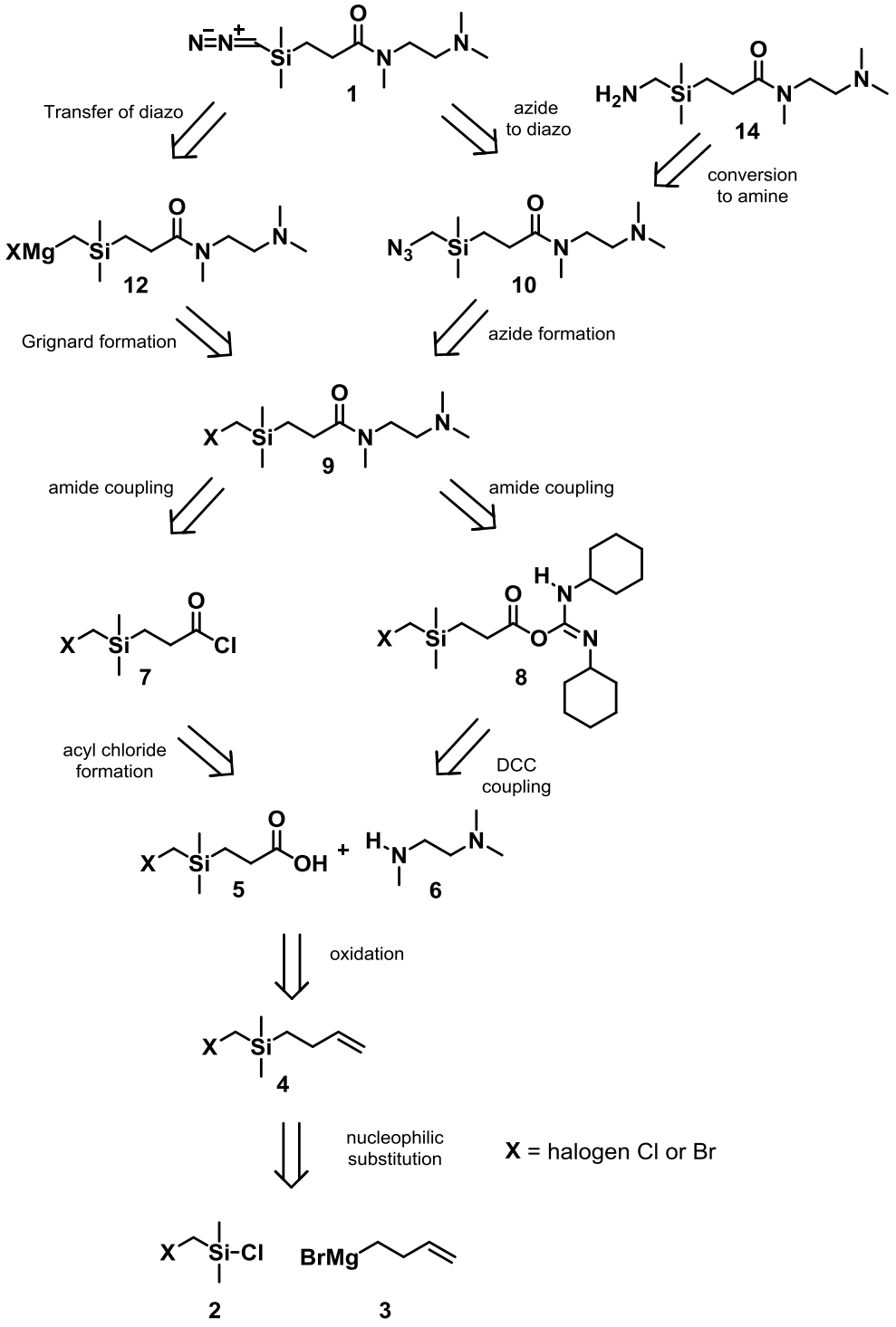


Figure 3-3 Design of water soluble carbamate trapping agent 1.

3-4 Retrosynthetic strategy for the chemical synthesis of designed carbamate trapping agent 1

A reterosynthetic approach was used to design a practical route towards target molecule **1** (Scheme 3-6).



Scheme 3-6 Retrosynthetic scheme for the synthesis of diazo trapping molecule 1.

To further the synthesis of molecule **1** the individual reaction steps need to be considered. From trapping molecule **1,** the first retrosynthetic step was to create the diazo group. The most common method used for TMS-DAM synthesis is the direct transfer of a diazo from diphenyl phosphorazidate (DPPA) to a carbanionic species (160). This reaction could occur via the formation of a Grignard on a halogen (**12**) which would then be replaced by the diazo from DPPA. This strategy requires a halogen to be in place attached to the silylmethyl group as seen in molecule **9**. As the work progressed the direct transfer of a diazo via this method proved difficult, and a further method to create a diazo through the interaction of a phosphine with an azide group was explored based on literature (142). Thus molecule **10** was also considered where the azide could be formed from a halide group via a Finkelstein reaction (161).

The design incorporates the tertiary amine functionality via an amide bond to introduce rigidity to the molecule, which is needed to prevent an intra molecular reaction. Amide coupling has been extensively covered in the literature (139). There are three common methods of amide coupling: 1) production of an acyl chloride, 2) via a mixed anhydride or 3) via an activated ester group. Using an ester introduces the additional step of ester activation so it was chosen to investigate the acyl chloride (via molecule **7**) and mixed anhydride (via molecule **8**) methods first. Both methods were found to be successful to some degree, but ultimately the acyl chloride method was taken forward due to higher yields and easier purification of the product from starting materials.

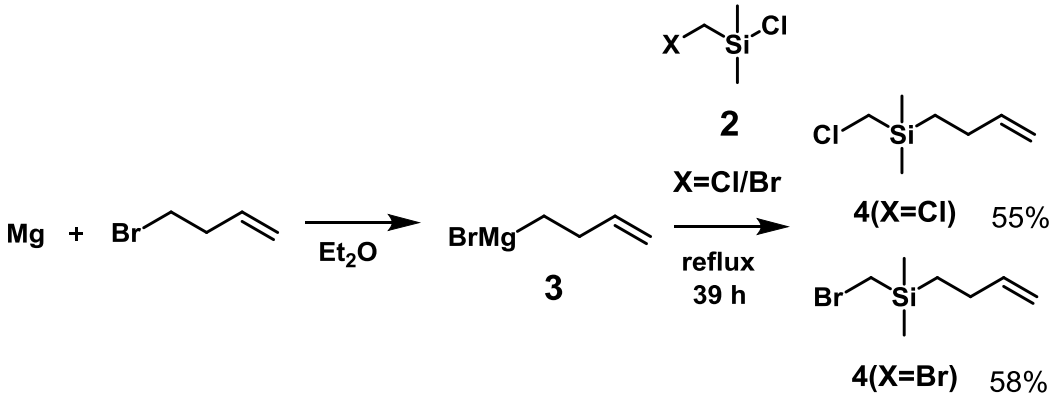
Both methods require a carboxylic acid starting material **5**. The disconnection of carboxylic acid molecules came from the literature (138). The procedure produced the carboxylic acid, **5**, by oxidisation of alkene **4**. Details of the development of the literature procedure are given later in this chapter. The production of the silyl alkene **4** was carried out following the same literature from the shown commercially available starting materials (**2**,**3**) via the production of a Grignard reagent and gave good yields.

The original starting point of the commercially available chlorosilane material **2** was chosen to match the backbone of TMS-DAM (Figure 3-2).

3-5 Forward Synthesis of Carbamate trapping agent 1

3-5.1 Synthesis of Alkenes

The first step was the synthesis of a silyl alkene **4**, either containing Cl or Br. These reactions have been carried out previously on both chlorine, **2** (**X**=Cl) and bromine, **2** (**X**=Br) chloro(dimethylsilyl)methylhalogen compounds (138,162). A but-3-enyl magnesium bromide Grignard reagent, **3,** was formed via the slow addition of 4-bromo-1-butene to a diethyl ether suspension of magnesium turnings. After the production of this Grignard reagent **3** the reaction proceeded via substitution of chloride from chloro(dimethylsilyl)methylhalogen, **2**, compound by the Grignard reagent to create (but-3-enyl)dimethylsilane halogen molecules **4** (Scheme 3-7).



Scheme 3-7 Synthesis of alkenes 4 from starting silyl material 2 and Grignard reagent 3.

The literature method used chlorosilane **2**(**X**=Cl), however, during the later stages of the project it was found that a more reactive halogen substituent was necessary, so the scope of the approach was broadened to encompass bromosilane **2**(**X**=Br).

The development of the reaction of the chlorosilane **2**(**X**=Cl) went through several phases (Table 3-1). Initially, there were difficulties in forming Grignard reagent **3**, most likely due to the presence of water. Precautions were taken to ensure an anhydrous environment, such as drying solvents and baking glassware. In addition, the introduction of iodine crystals served as a catalyst for Grignard reagent formation (163). The combination of these factors enhanced the formation of the desired haloalkenes **4**. Next, the development of the purification of the alkene

product was investigated. Isolation of the desired products from the silyl starting material (**2**) by vacuum distillation proved difficult due to the similarity in the boiling points. Therefore column chromatography was explored which greatly increased the yields to 36 % (Cl) and 48 % (Br) but these were still lower than the literature value of 53 % for the chlorosilane **2**(**X**=Cl). In order to improve yields, the use of an excess of Grignard reagent **3** was explored because, upon aqueous work-up, Grignard **3** will be protonated to yield 1-butane, which is gaseous, and readily removed. This approach removed the need for chromatographic purification, and produced the highest yields of halosilane products 55 % (Cl) and 58 % (Br) (Table 3-1).

Table 3-1 Details of optimisation experiments towards the preparation of haloderivatives 4 from Grignard reagent 3 and halosilanes 2 figure 3-6.

Attempt no.	**X**	**Ratio of 2 to 3**	**Yield/%**	**Comments**
1	Cl	1:1	0	No Grignard formed
2	Cl	1:1	12	Purified by distillation
3	Cl	1:1	36	Purified by column chromatography
4	Br	1:1	48	
5	Cl	1:2	55	No purification needed
6	Br	1:2	58	

The syntheses of the alkenes **4**(X=Cl) and **4**(X=Br) were confirmed by comparing ^{1}H NMR chemical shifts to values reported in the literature (138). In addition, comparisons of ^{1}H NMR signals between the starting materials 4-bromo-1-butene and the haloalkenes **4** were used. This is illustrated by the shifts of the signals c and d to c' and d' which are consistent with the conversion from 4-bromo-1-butene (**A**) to the bromosilylalkene molecule (**B**) (Figure 3-4).

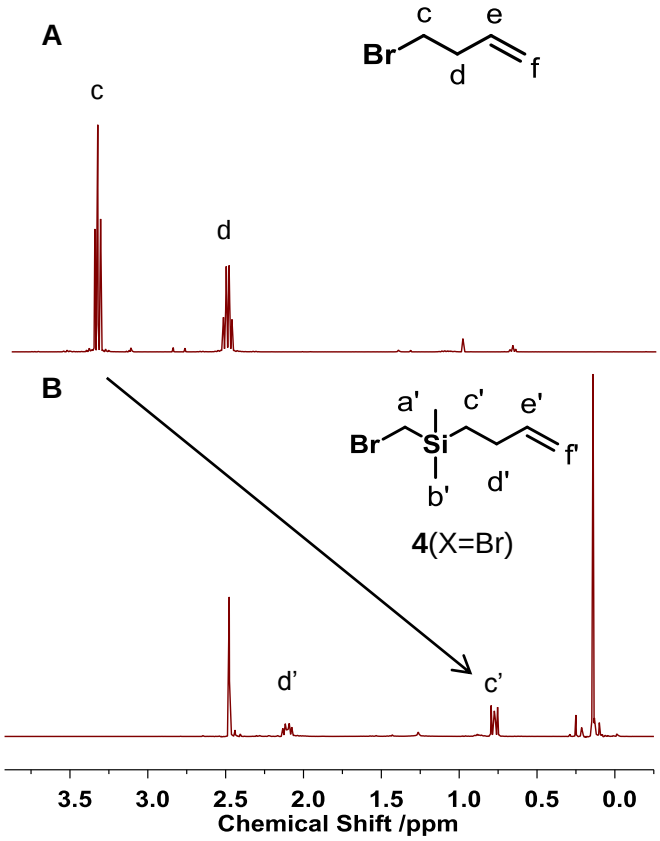


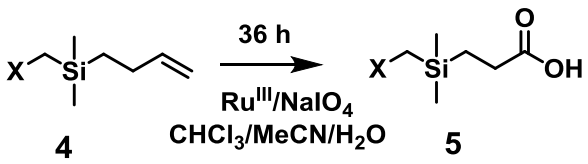
Figure 3-4 Comparison of 4-bromo-1-butene starting material (A) and halo alkene product 4(X=Br) (B).

At first, the Finkelstein reaction was used to convert the chloro alkene **2**(**X**=Cl) to a bromo alkene **2**(**X**=Br) before the azide reaction step but replacing the starting material with chloro(dimethylsilyl)bromomethyl, producing the final product **4**(**X**=Br) proved more efficient. (Details of the Finkelstein reaction are in section 3-5.5.)

3-5.2 Oxidation of alkenes 4 to acids 5

The next step of the synthetic strategy was the formation of carboxylic acids **5** by the oxidation of the alkene functionalities of haloalkenes **4** (Scheme 3-8). The process of oxidation of an alkene to an acid has been described many times and has been carried out on a similar system to haloalkenes **4** (138). However, the reported procedure used carbon tetrachloride as the solvent which is no longer used in laboratories for health and safety reasons. Therefore method development was

required to determine an alternative solvent system for this oxidation reaction. Chloroform was chosen as an alternative.



Scheme 3-8 Synthesis of carboxylic acids 5 by oxidation of haloalkenes 4.

The oxidation of haloalkene **4**(**X**=Cl) was initially carried out at the literature temperature of 23 °C but no reaction was observed. A second attempt was performed under reflux but this caused polymerisation of the starting alkene **4**. A mid-range temperature of 50 °C was then investigated, and this method was successful, but not over the literature time scale of 2 h. An extended reaction time of 36 h was then used, and this gave the desired product at reasonable conversion levels (**5**). Column chromatography was used and afforded isolated yields of **5**(**X**=Cl) (52 %) and **5**(**X**=Br) (77 %). Further optimisation using stepwise addition of the sodium periodate oxidant to regenerate RuO_4 (164), delivered almost complete oxidation of the bromo alkene **4**(**X**=Br) (165). These developments are summarised in Table 3-2.

Table 3-2 Details of method development for the oxidation of alkenes 4 to acids 5.

Attempt no.	**X**	**Temp. °C**	**Yield/%**	**Comments**
1	Cl	23	0	No observable reaction
2	Cl	80	0	Polymerised starting material
3	Cl	50	52	Column chromatography purification
4	Br	50	77	
5	Cl	50	74	Stepwise addition of sodium periodate
6	Br	50	98	

The formation of the acid group was monitored via IR-spectroscopy. The appearance of a signal at 1705 cm^{-1} corresponded to the formation of the carboxyl

group (166). The ^{1}H NMR spectrum also shows the disappearance of the vinyl signals from **4**(**X**=Br) (e and f) between 5.0 and 6.0 ppm (Figure 3-5) (and the shift of d to d' which is consistent with the expected shift value for a $RC\underline{H}_2CO_2H$ system).

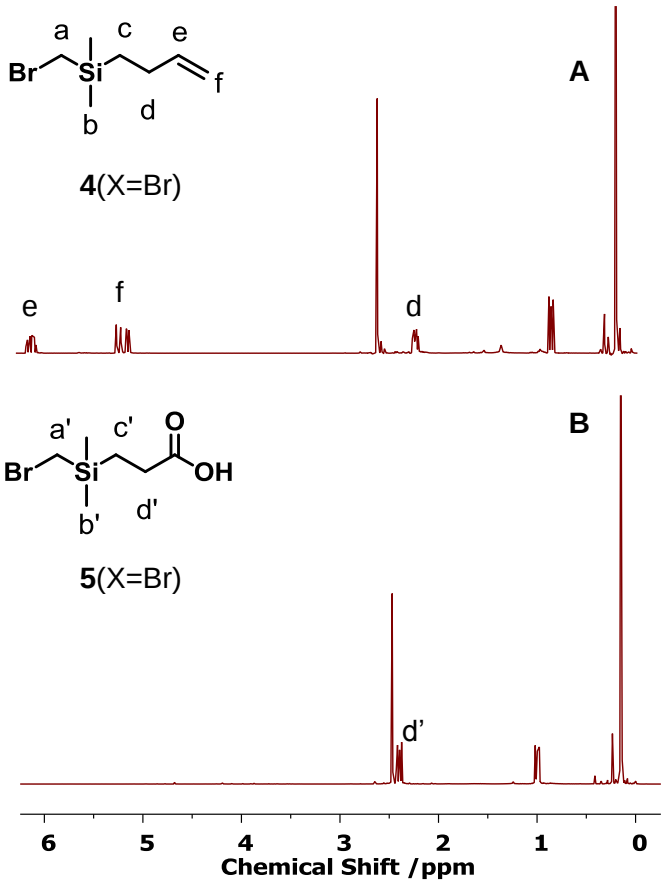


Figure 3-5 ^{1}H NMR comparison of the bromo alkene 4 (A) and Br acid 5 (B), showing the disappearance of the vinyl alkene signals (e and f) when the alkene (A) is oxidised to the carboxylic acid (B).

The carboxyl group is not seen on ^{1}H NMR spectroscopy however, the presence of C=O was confirmed by ^{13}C NMR signal at 179 ppm (e') (Figure 3-6). This signal is at a much higher chemical shift than the other molecule signals due to the high deshielding by the oxygen.

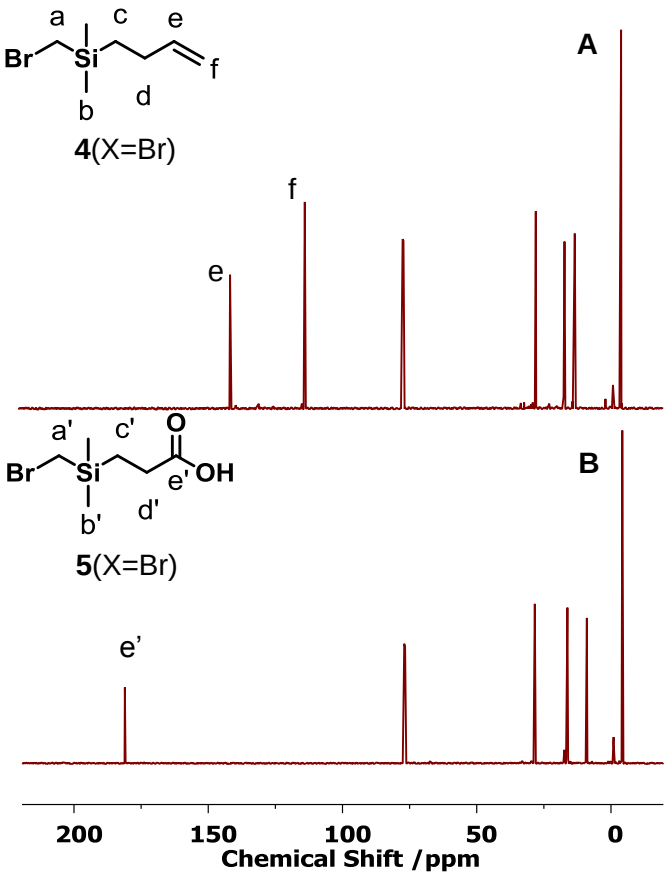
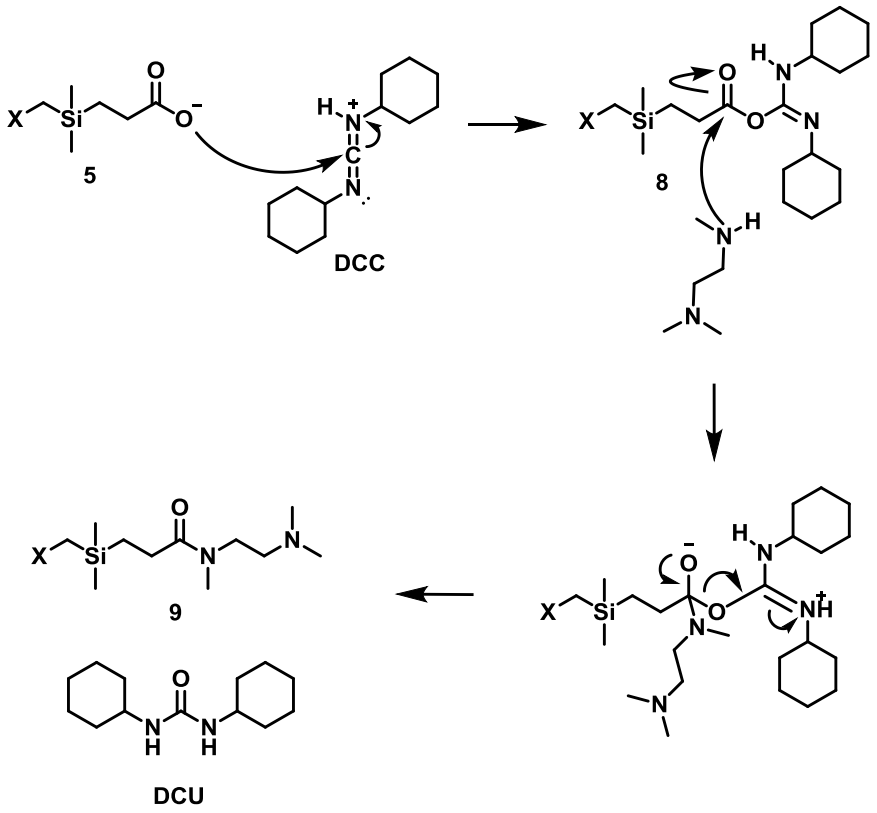


Figure 3-6 Comparison of ^{13}C NMR spectra of bromoalkene 4(X=Br) (A) and acid molecule 5(X=Br) (B) showing the appearance of the C=O signal at 179 ppm gives clear evidence for the formation of acid 5.

From haloacid compounds **5**, the water soluble element was introduced by the formation of amide groups on the newly synthesised carboxylic acid functionalities.

3-5.3 Formation of amides 9 from acids 5

There are several ways to form an amide bond from a carboxylic acid that have been reported in the literature (139), and two approaches were investigated. First a carbodiimide method, using the coupling agent dicyclohexylcarbodiimide (DCC) was explored. The DCC method works through the carboxylic acid reacting with the carbodiimide to produce an intermediate carboxylic ester (**8**) with a good leaving group. Displacement of this group by the addition of an amine gives the desired amide and the by-product dicyclohexylurea (DCU) (Scheme 3-9).



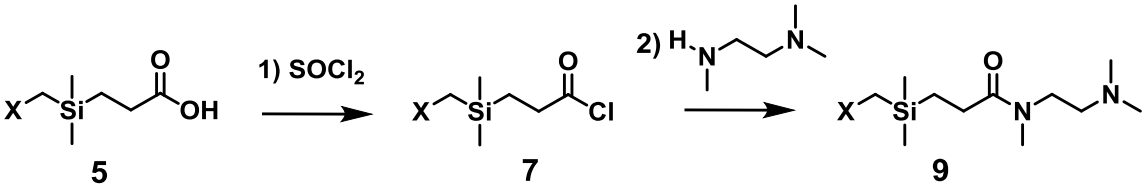
Scheme 3-9 Reaction mechanism of the formation of amides 9 from acids 5 using DCC.

The DCU by-product is sparingly soluble, and the majority of it can be filtered off after the reaction but removal of all traces proved difficult. Methods investigating the removal of DCU by centrifugation and cold ethyl acetate wash steps (167) gave impure materials with crude yields of 38 and 63 % respectively (Table 3-3). A different workup method involving the extraction of the amide into an aqueous layer by taking advantage of the charge state of the amine group gave effective purification, but with a very low yield of only 6 %. Because of these unsatisfactory results, the use of an acyl chloride intermediate was investigated for amide bond formation.

Table 3-3 Details of the development process of synthesis of amide 9 via DCC coupling method. Attempts 1 and 2 were never purified but produced crude yields of 38 and 63 % but attempt 3 produced purified product 6 %.

Attempt no.	X	Yield/%	Comment
1	Br	0(38)	Centrifugation
2	Br	0(63)	Cold ethyl acetate washes
3	Br	6	Basic workup

The second route investigated the formation of an acyl chloride. The method relies on the substitution of the hydroxyl group of the carboxylic acid by chlorine using thionyl chloride by chlorine that is in turn displaced by the amine group (Scheme 3-10). It proved important to keep the reaction under dry conditions as the acyl chloride group was readily hydrolysed back to the carboxylic acid **5** (168).



Scheme 3-10 Synthesis of amide 9 from carboxylic acids 5 via acyl chlorides 7.

Optimisation experiments were explored involving keeping the reaction dry, changing the temperature of the reaction and altering the ratio of the reagents as detailed in Table . The initial observation of no reaction was believed to be caused by water reacting with the acyl chloride. Thorough drying of the reaction solvent and then removal of the solvent altogether meant the reaction was able to progress to amine transfer. After the acyl chloride had been formed and reacted with amine **6** the main problem was the purification of the amides (**9**). The basic properties of the amides caused difficulties in purification via column chromatography due to their strong interaction with the acidic silica. As the amides **9** should be soluble in water when charged, a basic workup was developed to draw the starting material acid **5** into the water layer and leave the amide in the organic layer. Excess thionyl chloride decomposes on addition of water to give hydrogen chloride and sulphur dioxide which are both lost as gases at room temperature. So an excess of this reagent coupled with the basic workup was developed to produce purified product, with yields of 36 % **9**(**X**=Cl) and 97 % **9**(**X**=Br) (Table 3-4).

Table 3-4 Details of development of addition of amine 6 to acid 5 via an acyl chloride intermediate 7 to form amide 9.

Attempt no.	X	Yield/%	Comment
1	Cl	0	No reaction
2	Cl	<1	Loss during column chromatography
3	Cl	36	Repeated washes with basic workup and use of excess thionyl chloride
4	Br	97	

Confirmation of the formation of the amide was seen through the movement of the carbonyl IR stretch from 1705 to 1670 cm^{-1}. ^{1}H NMR spectra were compared to the spectrum of the amine starting material (**6**) which contains four signals in the range 2.0-3.0 ppm. Further corroboration came in the form of signal integration showing the correct number of protons and proton environments. The integrations of the signals also showed that there was no excess acid starting material remaining further confirming the effectiveness of the basic work-up as a purification strategy. These can be seen to correspond to **9**(**X**=Br) in Figure 3-7. Assignment by NMR spectroscopy is the best method for analysis of amide **9** because comparison to acid **5** will show very little movement of peaks due to very little change in the existing proton environments.

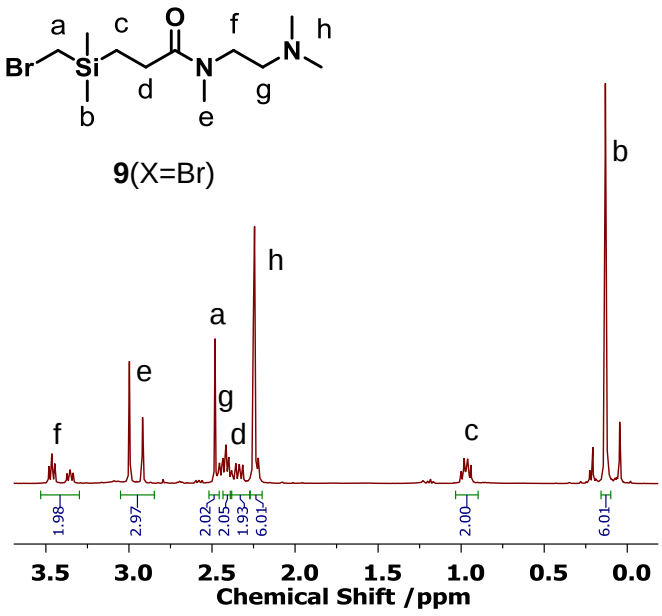


Figure 3-7 ^{1}H NMR spectrum for bromo amide 9.

Further confirmation can be seen through the movement of the carboxyl signal in the ^{13}C NMR spectrum when the amide bond forms from 181 ppm for the acid **5**(**X**=Br) to 173 ppm for amide **9**(**X**=Br).

The formation of an amide bond can take on two conformers, *cis* and *trans*. In Figure 3-7 signals e and f are seen as two signals due to restricted rotation around the amide bond, meaning that the product is present as both of these conformers (Figure 3-8).

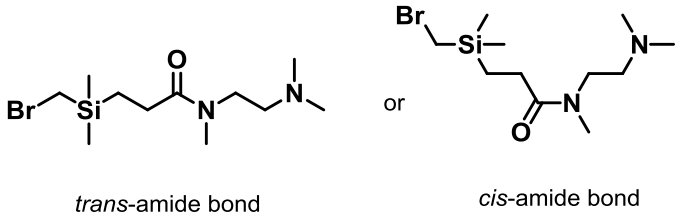
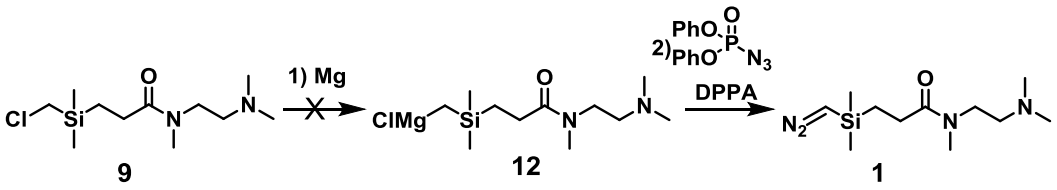


Figure 3-8 The *trans* and *cis* forms of the amide bond formation for amide 9.

3-5.4 Diazo group transfer to halo amides 9

Following the synthesis of halo amides (**9**), there were several different pathways to investigate the formation of the final diazo compound, **1**. In the literature the formation of TMS-DAM was usually by the transfer of a diazo group from an azide containing compound, most commonly DPPA (141,160). This reaction would proceed via the creation of a Grignard reagent (**12**). Based on this literature the first method to introduce a diazo group on **9** was via Grignard formation and diazo transfer (Scheme 3-11).



Scheme 3-11 Attempted synthesis of diazo molecule 1 via a Grignard reagent 12 and diazo transfer from DPPA.

This reaction never reached completion. This was due to failure to form the Grignard reagent **12** though several methods were trialled (Table 3-5). The lack of production

of Grignard **12** on the amide **9**(**X**=Cl) was attributed to the presence of water in the reaction. Commercial dry solvents and thorough drying of glassware were used but still no reaction between the chloride **9**(**X**=Cl) and magnesium was observed. The use of 1,2-dibromoethane as a Grignard initiation catalyst was then introduced (169) without success. It was at this stage that the use of bromoamide **9**(**X**=Br) was investigated to increase the reaction of the halogen group and improve Grignard formation. The alteration of the chloride for a bromide group was carried out with use of the Finklestein reaction (3-5.5). The Grignard reaction on **9**(**X**=Br) was also not successful.

Table 3-5 Details of attempts towards diazo group transfer to amide 9.

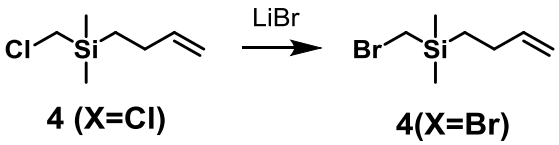
Attempt no.	X	Yield/%	Comment
1	Cl	0	Reaction not dry enough
2	Cl	0	Addition of dibromoethane catalyst
3	Br	0	Tried a more reactive halogen
4	Br	0	Purchased dry solvents rather than drying with molecular sieves

There are several other methods of diazo formation that have been explored within the literature so another experimental approach was investigated. Work by Myers and Raines, (2009) (see sections 3-5.6 and 3-5.7) investigated the conversion of an azide group to a diazo. In order to examine this strategy the halide **9** was converted to azide **10**. The chemistry of exchanging a halogen to an azide is very well established and has been carried out before (140) on different systems.

3-5.5 Interconversion of halo substituents of alkenes 4 via the Finkelstein reaction

The Finkelstein reaction is used to interconvert halogens (161). This reaction was used to increase the reactivity of the halomethylene (XCH_2) group after failure to produce a Grignard on the chloride derivative of molecule **9**. The chlorine was

substituted with both bromine (Scheme 3-12) and iodine in attempts to increase the halogen reactivity for the production of the Grignard reagent (**12**) needed for diazo transfer.



4 (X=Cl) **4(X=Br)**

Scheme 3-12 Synthesis of bromo alkene 4 (X=Br) by Finkelstein reaction on chloro alkene 4(X=Cl).

The transfer of the bromine in place of the chlorine was easily monitored with ^{1}H NMR by the movement of the proton peak connected to the halogen (a) but with a maximum yield of 64 % (Figure 3-9).

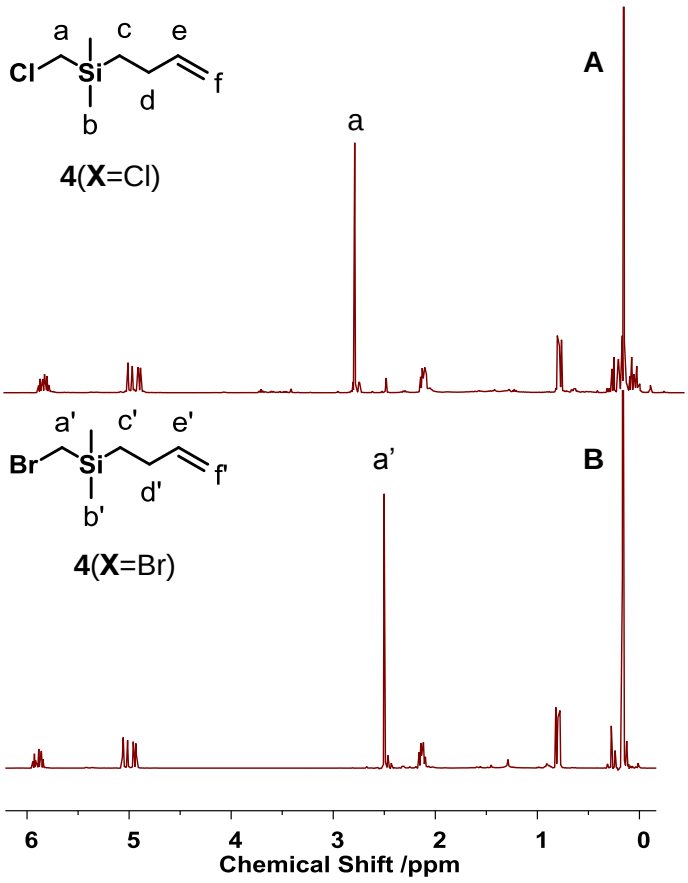


Figure 3-9 ^{1}H NMR showing the movement of signal a from 4(X=Cl) (A) to a' 4(X=Br) (B).

At first the yield seen was relatively low (5 %) this was probably because the reaction time was too short for this system. The yield was then increased by the use of dry acetone to reduce any possible hydration of the halogen group (170). Besides hydration there are no other side reactions taking place so the reaction time was increased to 40 h which increased the yield to 37 %. After this, the sodium bromide (NaBr) reagent was replaced with lithium bromide (LiBr) which is more soluble in acetone and this alteration increased the yield to 64 % (Table 3-6).

Table 3-6 Details of the optimism of the Finkelstein reaction to convert alkene 4 (X=Cl) to alkene 4 (X=Br).

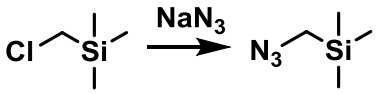
Attempt no.	Halogen converted	Time /h	Yield /%	Comment
1	Cl-Br	2	5	Reaction time too short
2	Cl-Br	4	15	Dry acetone used
3	Cl-Br	40	37	40 h
4	Cl-Br	24	64	Exchanged reagent

Chlorine was substituted for iodine in the same way with a yield of 50 % but the iodine group on the alkene proved to be too reactive during the oxidation of iodoalkene **4**(**X**=I) to iodoacid step cyclisation was observed. For this reason, bromine was chosen as the best halogen group to continue development of the overall synthesis.

3-5.6 Halide to azide conversion

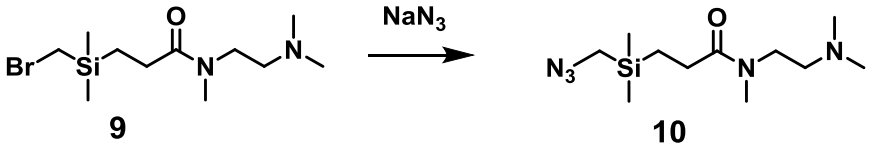
There has been extensive work done on the conversion of azide groups to diazo groups across a range of compounds (142). Reactions were carried out in aqueous solution using phosphine compounds. In order to attempt this chemistry, the halogen group of haloamides needed to be converted into an azide group. This reaction has

previously been carried out on the trimethylsilylmethylchloro compound (Scheme 3-13).



Scheme 3-13 Literature method of converting a chlorine to an azide group.

This previous work on a similar silyl backbone was successful so these conditions (heating in dried DMF with sodium azide) were the starting point for the method development on molecule **9**(**X**=Br) (171) (Scheme 3-14).



Scheme 3-14 Synthesis of silyl azide 10 from silyl bromo amide 9 via nucleophilic substitution with sodium azide.

For this step the use of bromide **9**(**X**=Br) instead of chloride **9**(**X**=Cl) as the halogen moiety became important due to the increased reactivity of the bromide providing a better leaving group. When using **9**(**X**=Cl), complete consumption of starting material **9** was never observed. The development of this reaction involved increased reaction time. The original reaction temperature of 80 °C gave only a 60 % conversion of the starting molecule **9** to product **10** (seen by ^{1}H NMR). By reducing the temperature to RT and experimenting with the reaction time using 24 h and 48 h, the yield increased to 84 % and 92 % respectively (Table 3-7).

Table 3-7 Details of development process from synthesis reaction step Scheme 3-14.

Attempt no.	X	Yield/%	Comment
1	Cl	0	Trace levels
2	Br	0	60% conversion by ^{1}H NMR
3	Br	84	24 h
4	Br	92	48 h

The reaction was monitored by the appearance of an azide IR signal at 2091 cm^{-1} which is a very distinct signal and by the movement of signal a to a' using ^{1}H NMR spectroscopy from **9**(**X**=Br) to azide **10** (Figure 3-10).

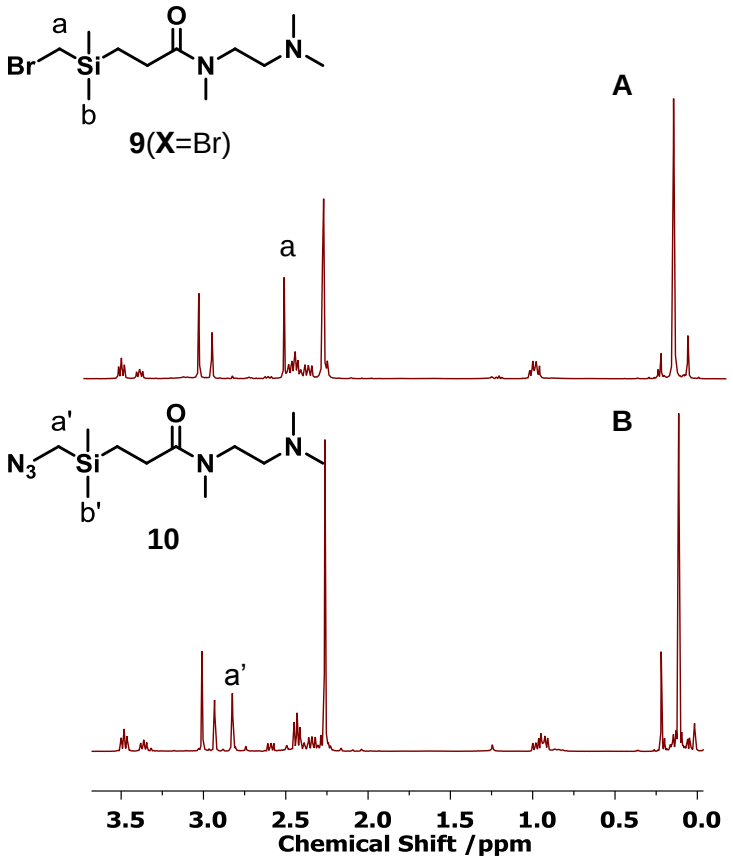
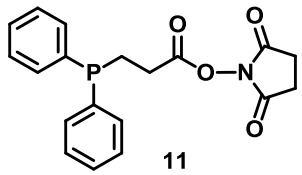


Figure 3-10 ^{1}H NMR comparison of bromo amide 9 (A) to spectra for azide amide molecule 10 (B).

3-5.7 Conversion of azide 10 to diazo 1

As previously mentioned, Myers *et.al.* 2009 studied extensively the conversion of azide groups to diazo groups. The reaction procedure involves the addition of a phosphine to the azide group which is then hydrolysed to form a diazo. The literature described several commercially available phosphine groups, the phosphine selected for this azide to diazo conversion had been previously used on an RCH_2N_3 group considered to be chemically similar to molecule **10** (Figure 3-11).



11

Figure 3-11 Phosphine 11

Figure 3-12 shows the proposed mechanistic steps for the attachment and subsequent loss of the phosphine (**11**) adapted from Myers *et.al.* 2009 (142) which is the fragmentation of 1,3-disubstituted alkyl aryl triazenes. This reaction never reached completion but instead was halted at the fragmentation of the bound phosphine to the azide group (**13**).

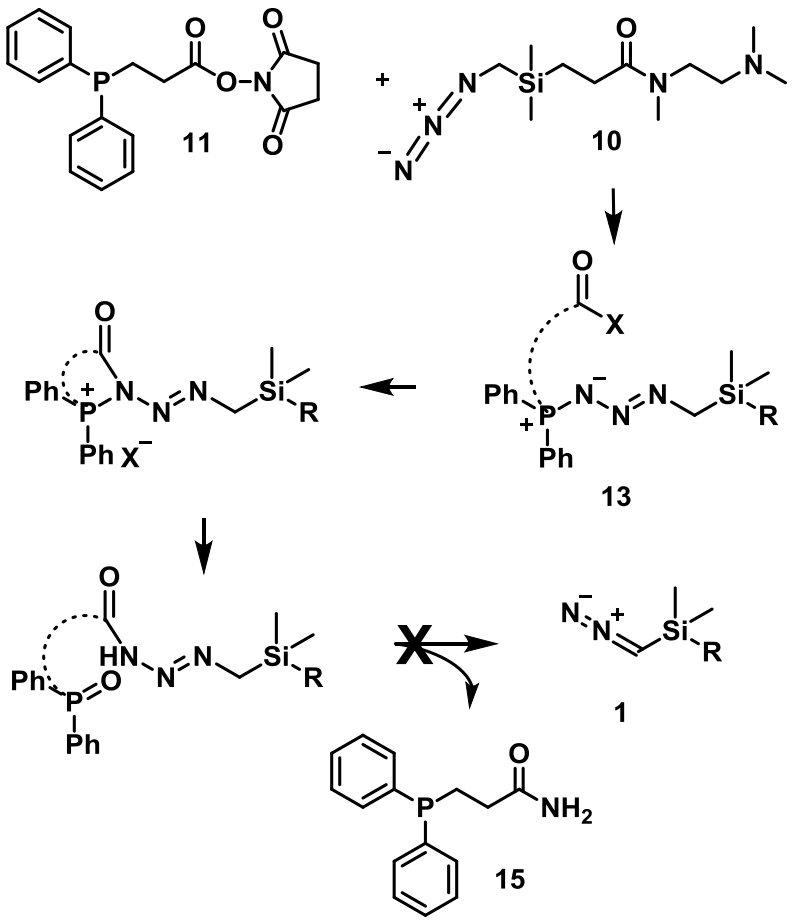


Figure 3-12 Suggested reaction mechanism between a phosphine molecule binding to the azide functional group and being displaced to form a diazo.

^{1}H NMR and IR spectroscopies confirmed the absence of the methyleneazide signal but did not give evidence for the diazo group, however the NMR spectrum contained many by-product signals which made analysis difficult. For this reason, ^{31}P NMR spectroscopy was used. It was confirmed that phosphine **11** was being converted into the intermediate product of an acyl triazene (**13**) by a shift from −15 ppm to 30 ppm suggesting that the phosphine was successfully coupled to the azide but failed to proceed through to diazo formation. One difficulty analysing this reaction with ^{31}P NMR spectroscopy was that the intermediate and final product both have very similar signals (acyl triazene 30 ppm, released phosphine (**15**) 36 ppm) (142) and the shift pattern of a ^{31}P NMR is more difficult for comparison between samples due to the lack of a reference signal in the solvent.

From this intermediate step, several different workup conditions were investigated

in order to facilitate conversion of the acyl triazene intermediate to diazo product (**1**) (Table 3-8). The workup described in the literature for the use of phosphine **11** was investigated first. The procedure involved a basic workup using $NaHCO_3$, however this was shown to be unsuccessful as the ^{31}P NMR spectroscopy showed no change in the assumed acyl triazene signal (~30 ppm), the assumed reason was that the phosphine had bound more tightly to molecule **10** than the one used in literature. It is possible that the silyl group within molecule **10** could be stabilising the acyl triazene. Next a wash with 10 % NaOH was used. This strategy was adapted from the wash system of diazo transfer from a tosy or mesyl azide system (172) but the phosphine remained attached. Other conditions of phosphate buffer and the use of an acidic wash were then used to see if these conditions could fragment the phosphine but all proved unsuccessful in fragmenting the acyl triazene **13**.

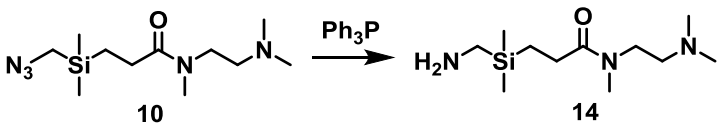
Table 3-8 Summary of workup conditions and outcomes for the attempted breakdown of intermediate acyl triazene to diazo 1 (conversion % calculated by ^{31}P NMR).

Attempt no.	Reaction conditions	Conversion of phosphine starting material (11) to intermediate (aryl triazene) /%	Workup
1	Dry, 2 h	0	Basic ($NaHCO_3$)
2	Dry, 12 h	0	Basic (NaOH)
3	H_2O, 12 h	40	Phosphate buffer
4	H_2O, 24 h	90	Acidic

None of the workup methods removed the phosphine attached to the silyl molecule to form the diazo target **1**. In light of this, a third method of converting an amine into a diazo group was found in the literature (173) and this is summarised below. This approach required the conversion of the existing azide into an amine group (Scheme 3-15).

3-5.8 Conversion of azide 10 to amine 14 for conversion to diazo 1

Initially the Staudinger reduction of azide **10** using PPh_3 was explored (Scheme 3-15).



Scheme 3-15 Synthesis of silyl amine (14) by reduction of azide (10) with triphenylphosphine.

The conversion of azide to amine was successfully carried out using a literature procedure with the use of triphenylphosphine (174) monitored by IR. Literature work described the conversion of the amine group to a diazo using nitrous acid (175) or the use of triflyl azide (173) however due to time constraints these methods were not investigated.

3-6 Conclusion

This chapter discussed the development and partial synthesis of trapping molecule **1**. Unfortunately, successful synthesis of the diazo group was not achieved despite the multiple methods investigated (direct diazo transfer and conversion from an azide). This work was carried out in parallel with the development of a method using a commercially available reagent for trapping (Meerwein's reagent). Ultimately the use of Meerwein's reagent proved successful for the trapping of carbamates (discussed in Chapter 4). Therefore the use of Meerwein's reagent took precedence over efforts towards the synthesis of diazo target **1**. Though the diazo was never formed, the preceding steps were all optimised into robust procedures. There are still several methods of creating a diazo group that have not been investigated, and it is likely that this work will be continued by another member of the research group in the future. A summary of these methods is provided in Future Work (section 3-7).

3-7 Future work

This synthetic route has been developed to the final stage. The original method of diazo group introduction in this investigation was via the transfer of DPPA as this is the method most often used in the literature. There are three key ways the method could be improved. First, methods to activate the magnesium could increase the chances of Grignard formation. Magnesium can be activated by dry stirring in an inert atmosphere (removing the oxide from the magnesium surface) or the use of the catalyst vitride (removes moisture from the metal surface) (169). Secondly, the Finklestein reaction could be carried out on amide **9** to introduce the iodine at this later stage and create a more active leaving group for Grignard formation. Finally, a different diazo transfer reagent could be used for the same reaction mechanism; the use of tosyl azide instead of DPPA has been demonstrated in several cases for the synthesis of a diazo group (176).

The work involving the use of the phosphine (**11**) for diazo formation could also be re-visited to try to remove the attached phosphine group with harsher conditions and longer reaction times. The previous work in this area described the synthetic use of several different phosphine groups and another could be used in place of **11** to improve this approach.

There is also another method of diazo formation that was not investigated. This is the reaction of an acyl chloride with diazomethane to form a diazoketone using the Arndt-Eistert reaction (177). This method would involve converting the halogen group into a carboxylic acid and then reacting with thionyl chloride though these reactions might be too harsh for the rest of the molecule.

Chapter 4: Development of a carbamate trapping method

4-1 Overview

As previously mentioned (chapter 3) this investigation discusses the development of two methods for the trapping of a carbamate within a protein. Chapter 3 explained the synthetic approach carried out towards synthesising a water-soluble TMS-DAM derivative to achieve this objective. The work in this chapter explores the second approach, using a commercially available alkylating reagent. The reagent discussed in this chapter is a salt known as Meerwein's reagent; a term covering both the trimethyloxonium tetrafluoroborate (TMO) and triethyloxonium tetrafluoroborate (TEO) derivatives (Figure 4-1).

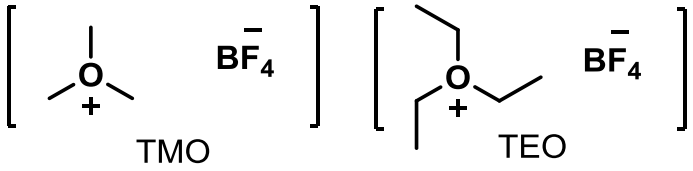


Figure 4-1 Meerwein's reagents trimethyloxonium tetrafluoroborate (TMO) and triethyloxonium tetrafluoroborate (TEO).

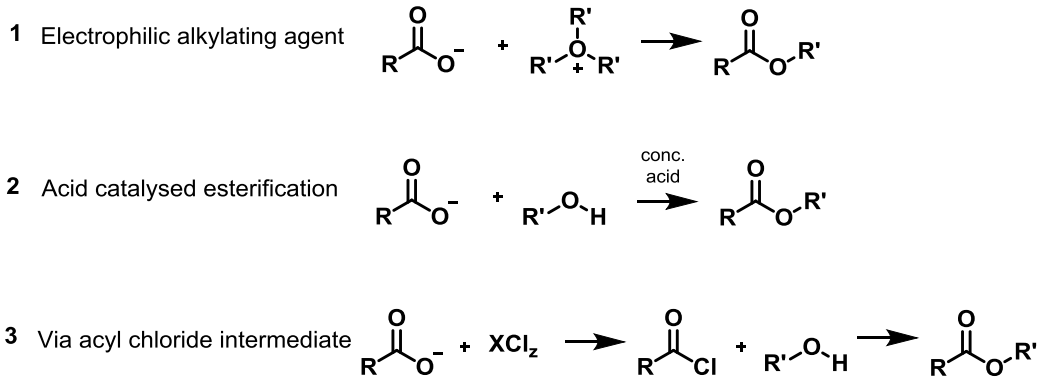
The basis for this trapping experiment follows the same principles as the use of TMS-DAM; the transfer of a methyl or ethyl group to the formed carbamate to create a more robust sample for downstream analysis.

The results described within this chapter demonstrate the development of a method in which Meerwein's reagent can be used to trap carbamates under physiologically relevant conditions. This method was begun using TMO but was exchanged for TEO due to the fast hydrolysis of TMO within an aqueous system. Method development began using single amino acids and dipeptides before successfully trapping CO_2 on a protein already known to form a carbamate, haemoglobin (Hb).

4-2 Methods to modify carboxyl groups and previous work with Meerwein's reagent

The main challenge of studying carbamates within a protein has been a lack of tools to investigate this modification under physiological conditions due to its labile nature. This work aims to modify the carbamate by the addition of a covalently bound group to make the carbamate robust enough for downstream analysis by MS.

Previous trapping of carbamic acid species in a mostly organic solvent system (methanol:benzene, 1:4) was carried out by methyl transfer from TMS-DAM (116). This is the process of alkylation via an electrophilic reagent. Several methods of alkylation on a carboxyl group have been investigated in the literature (Scheme 4-1); alkylation with an electrophilic reagent (**1**) such as Meerwein's or TMS-DAM, esterification using an acid catalyst (**2**) and group transfer via an acyl chloride intermediate (**3**).



Scheme 4-1 Modification of a carboxyl group by three different alkylation methods, electrophilic reagent (1), acid catalysis (2) and via an acyl chloride (3) to form an ester.

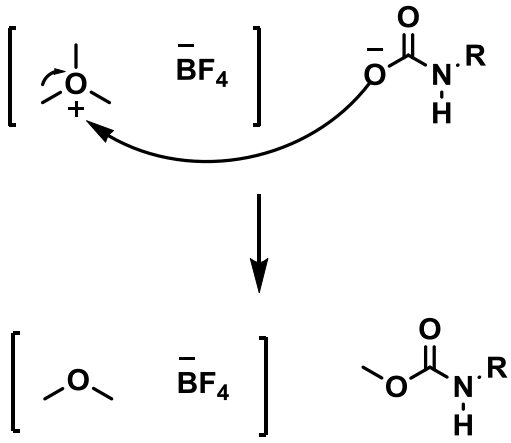
Meerwein's reagents have been previously investigated in comparison to these other alkylation methods. It was concluded that esterifying with an electrophilic reagent is the best direct method with no problem of reversibility or need for any hazardous chemicals (139). Acid catalysed esterification requires a strong acid to activate the carboxyl and the alcohol; the addition of a strong acid would perturb the cellular conditions that this investigation is trying to maintain (170). The method using an acyl chloride intermediate involves the use of a harsh chemical to produce

an intermediate which would also disturb a cellular environment (171). Therefore, the use of an electrophilic alkylating agent is the best choice for modification of a carbamate while limiting any disturbance to the physiological conditions being mimicked.

Whilst there are several electrophilic alkylating agents to choose from - Meerwein's reagents are most applicable for this work as they are tetrafluoroborate salts and therefore water-soluble. They are also much safer than the alternatives such as methyl iodine or dimethyl sulphate (172).

4-2.1 Trimethyloxonium tetrafluoroborate (TMO)

Previous work with TMO has converted organic acids into methyl esters as a safer alternative to diazomethane (139,173,174). TMO has also been reacted with several different proteins to modify the carboxyl groups present to investigate enzyme activity. The process of how TMO would modify a carbamate is shown in Scheme 4-2.



Scheme 4-2 Hypothesised reaction mechanism of methyl transfer from TMO to a carbamate group within a protein.

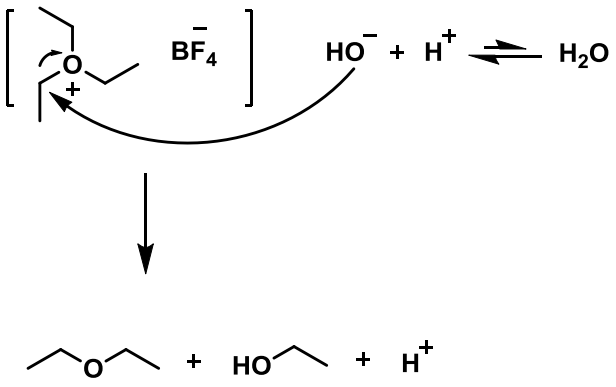
There have only been a few experiments carried out with TMO in an aqueous system and they required vigorous stirring and incremental addition of TMO to

counter the hydrolysis of the reagent (175). The modification of catalytic carboxyl groups in Pepsin was carried out in aqueous solution using TMO (176) and showed an activity decrease with increased transfer of methyl groups (methyl group transfer was measured using ^{14}C labelled TMO (176)). However, this work by Paterson and Knowles (1972), does not look into the location of the transferred methyl groups or what other affects this reagent might have on the protein but simply reports the correlation between increased methylation and reduced protein activity. The use of TMO to reduce activity in Ca^{2+} activated K^+ channels was measured by a reduction in current after TMO addition (175). This previous work demonstrates that TMO has been shown to modify proteins under aqueous conditions (174).

4-2.2 Triethyloxonium tetrafluoroborate (TEO)

The main reactive difference between TMO and TEO is the rate of hydrolysis in aqueous solution. TMO has a half-life in aqueous solution of 0.14 s (175) whereas TEO has a half-life of 7.4 min (117). The investigation described in this chapter began with the use of TMO but moved to TEO when it became apparent the longer half-life was necessary.

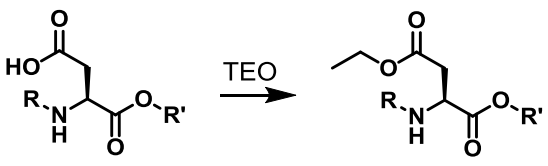
Another difficulty associated with the use of these reagents is the pH change caused by their hydrolysis, which forms H^+ ions reducing the pH of the solution (Scheme 4-3). This was important within this work because in order to accurately replicate a cellular environment the pH must be carefully maintained at 7.4.



Scheme 4-3 Hydrolysis of TEO reagent producing H^+ ions which causes a drop in pH of the solution.

Previous experiments using this reagent have used the manual addition of a base (normally sodium hydroxide) to maintain the desired pH (117). The development of the work recorded here began in this way but it was clear that the range of pH created by this method was too wide to accurately represent a cellular system. A pH stat was therefore introduced to maintain pH with the slow addition of 1 M sodium hydroxide (NaOH) during the reaction. TEO's advantage over TMO is that the slower hydrolysis rate improves the ability to buffer the system around a chosen pH and allows for a constant environment during the trapping reaction.

The use of TEO in aqueous solutions has previously been investigated as an ethylating reagent (117). To counter the hydrolysis of the reagent a 10-20 times molar excess was used and the esterification of organic acids was successfully confirmed by ^{1}H NMR spectroscopy. The use of TEO to modify carboxyl groups was used to reduce the enzymatic activity of lysozyme by carboxyl esterification (177). A generic example of ethylation of an aspartate by TEO is demonstrated in Scheme 4-4.



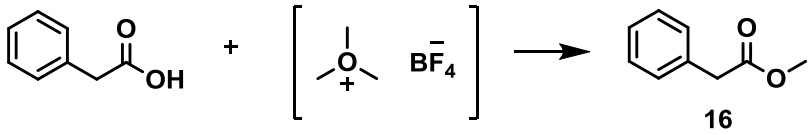
Scheme 4-4 Transfer of an ethyl group from TEO to the carboxylic acid side chain of amino acid aspartate.

As with the previous TMO investigation, no further analysis beyond the activity of the enzyme was established. This means that firm conclusions cannot be drawn about the sites that were modified or which of those were important for enzyme activity. It is possible that the most likely reason for the activity loss was the total denaturation of the enzyme.

4-3 Carbamate trapping using TMO

4-3.1 Phenylacetic acid

The first step of this investigation was to explore whether TMO could transfer a methyl group to a carboxylic acid within a physiologically relevant environment. To achieve this the TMO reagent was tested on an organic molecule, phenylacetic acid (2-2.1.1). This test was modelled on a cellular environment using a phosphate buffer at pH 7.4. Phosphate buffer was chosen as it buffers in the physiologically relevant pH range and does not contain any amine groups which might interfere with the trapping reagent. It was hypothesised that the reagent would transfer a methyl group to the carboxyl group of the acid producing phenylacetate (**16**) (Scheme 4-5).



Scheme 4-5 Transfer of a methyl group from TMO to the carboxyl group of phenylacetic acid to produce phenylacetate.

Manual addition of NaOH was used to maintain the pH around 7.4. Due to the short half-life of TMO in water, a lot of the reagent hydrolysed over the reaction so a 20 times molar excess of TMO was used (chapter 2 2-2.1.1). The results of this reaction were analysed by NMR spectroscopy (Figure 4-2).

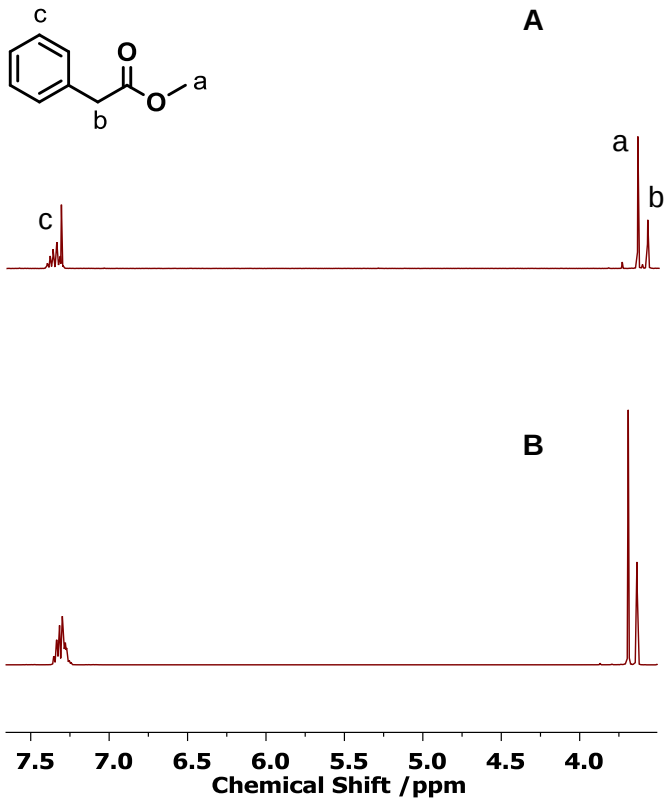


Figure 4-2 ^{1}H NMR comparing synthesised phenylacetate 16 from the reaction of TMO and phenylacetic acid (A) and purchased phenylacetate from Sigma Aldrich (B).

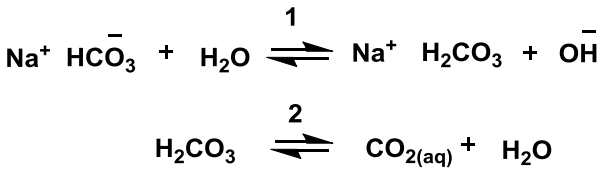
The synthesised phenylacetate was produced with a 93 % yield. The appearance of a methyl signal at 3.69 p pm using ^{1}H NMR (a) confirmed the presence of the synthesised phenylacetate product which corresponds to commercial phenylacetate material.

A mass peak of 150.0 m/z seen by ESI-MS also confirmed the product. These experiments demonstrate that despite hydrolysis, TMO is able to transfer a methyl group to a carboxyl in an aqueous environment.

4-3.2 Acetyl-lysine

Once the reagent was shown to carry out methyl group transfer, the next step was to establish whether a carbamate could be trapped under physiological conditions. Acetyl-lysine was used because the ε-amino group of the lysine side chain is known to form carbamates *in vivo* (94). Before trapping, the formation of a carbamate under the current experimental conditions was investigated to demonstrate carbamate formation.

Carbon dioxide was introduced to the system using sodium bicarbonate ($NaHCO_3$). In solution $NaHCO_3$ dissociates to form carbonic acid (**1**), which then dissociates to give carbon dioxide (**2**) (Scheme 4-6).



Scheme 4-6 Dissociation of sodium bicarbonate to carbonic acid which then produces carbon dioxide.

The formation of the carbamate was analysed using ^{13}C NMR spectroscopy (20 mM $NaHCO_3$) (Figure 4-3).

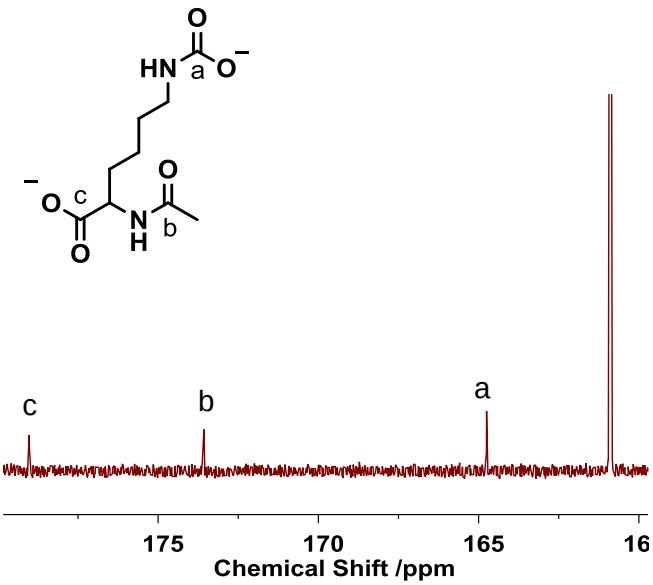
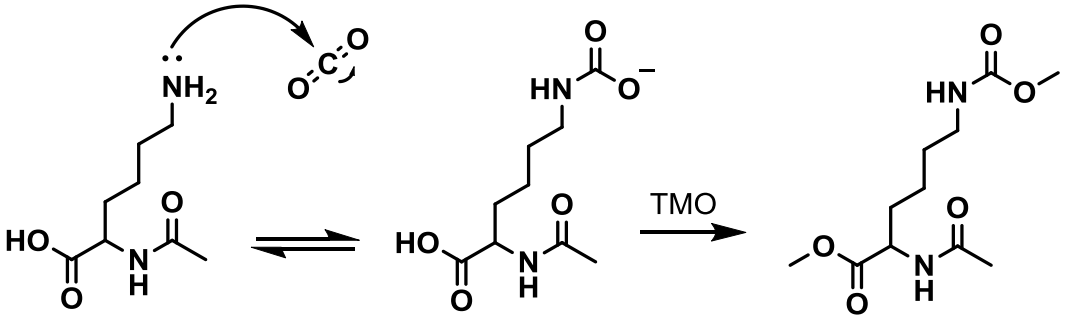


Figure 4-3 ^{13}C NMR spectrum showing the formation of a carbamate on acetyl-lysine by the appearance of a peak at 164 ppm (a). (Extra peak at 161 ppm is bicarbonate in solution.)

The carbamate formation was confirmed by the appearance of a peak at 164 ppm matching literature values (137).

After confirmation that the carbamate was forming, the TMO reagent was used with the developed conditions to test whether a methyl group could be transferred to this formed carbamate. $NaHCO_3$ was used at a 20 mM concentration as a physiologically relevant concentration for animal systems (178). The manual addition of NaOH was used to keep the pH at approximately 7.4. It was hypothesised that a methyl group would be transferred to the carboxyl group of the acetyl-lysine as well as to the formed carbamate (Scheme 4-7).



Scheme 4-7 The process of carbamate formation on the ε-amino group of the lysine and the predicted results of the transfer of a methyl group from TMO to both the carbamate and the carboxyl group.

The reaction was analysed using electrospray mass spectrometry (ESI-MS) (Figure 4-4).

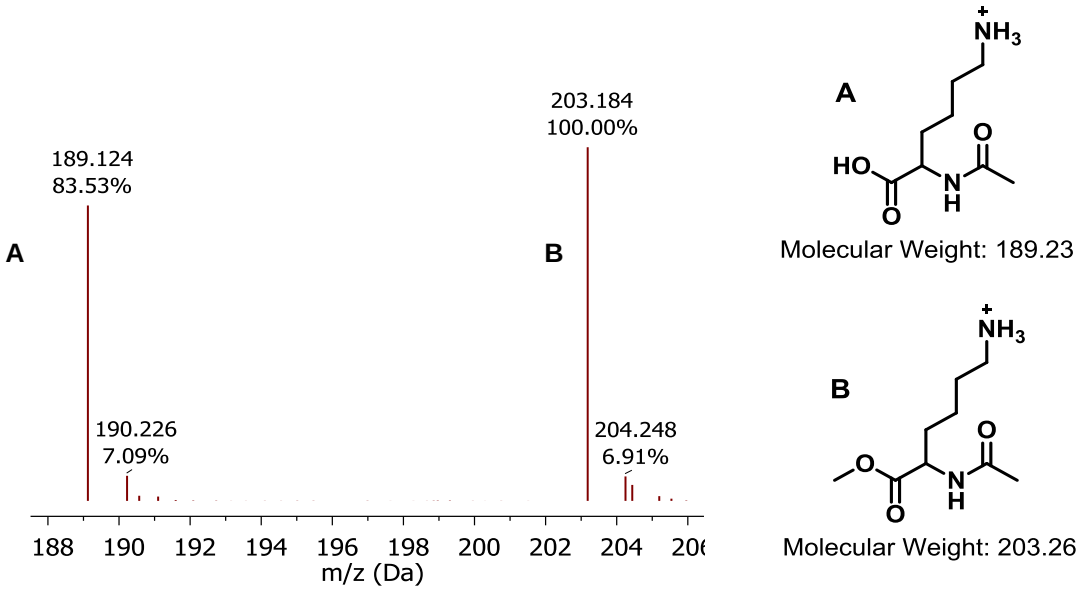


Figure 4-4 ESI-MS spectrum showing acetyl-lysine starting material (A) and acetyl-lysine with a methylated carboxyl group (B).

Unfortunately, no methyl trapped carbamate product was found, the ESI-MS trace showed only starting material acetyl-lysine (**A**) and methylated acetyl-lysine (**B**).

Methyl transfer onto the carboxyl group was expected from the previous experiment with phenylacetic acid. However, there was also a very small peak in the chromatogram (not shown) equivalent to the mass of acetyl-lysine with two additional methyl groups, one to the carboxyl and one to the amine group (Figure 4-5).

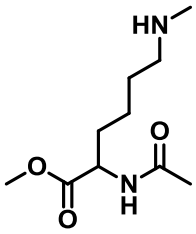
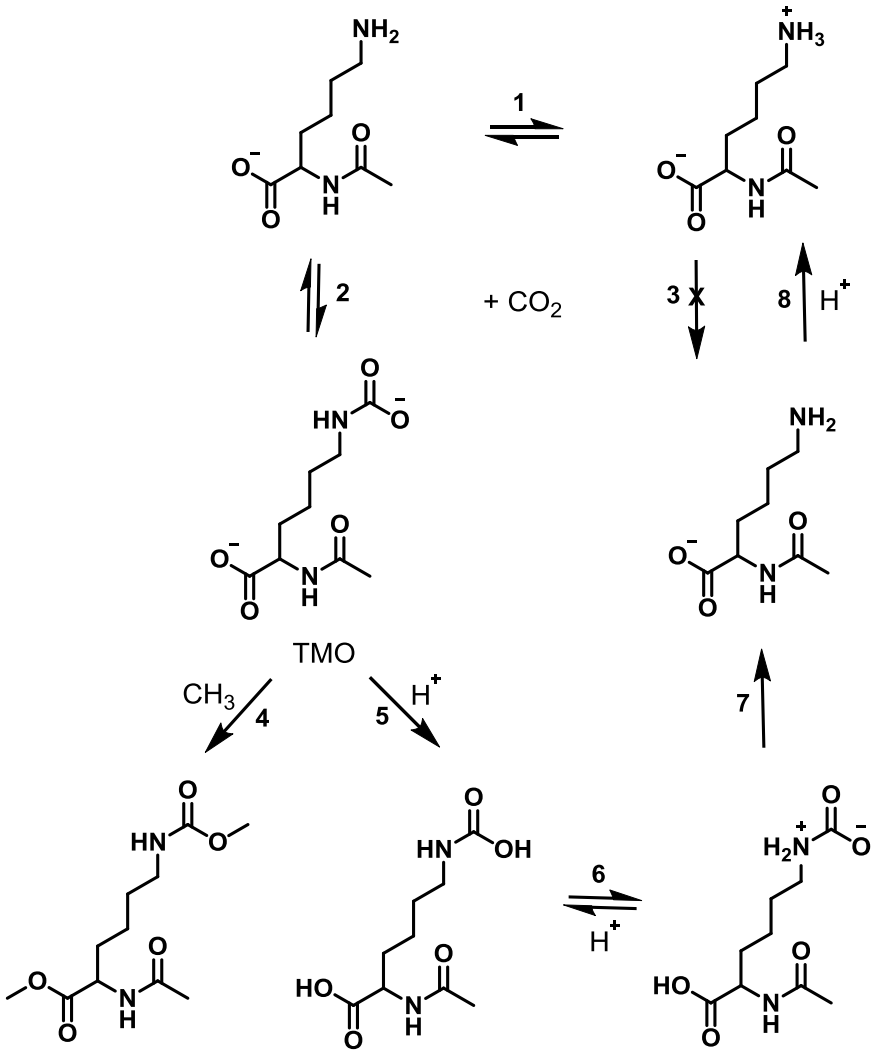


Figure 4-5 Acetyl-lysine with additional methyl group transfer on the carboxylic acid and amine side chain.

Though the reagent is more likely to transfer to the carboxyl than the amine, the excess of reagent needed to counter the hydrolysis reaction means that transfer onto an amine is possible. This was not expected to be a problem as carbamate formation only occurs on uncharged amines. Nonetheless, these two side reactions need to be accounted for when carrying out whole protein analysis.

The process of carbamate formation within this system involves several equilibria. An overview of these is shown in Scheme 4-8. Under the pH of the reaction the amine group of acetyl-lysine is in equilibrium between its charged and uncharged states (**1**). When CO_2 is added to the system, with the use of sodium bicarbonate, it will react with the uncharged amine group to form a carbamate (**2**) but no reaction will occur with the charged amine (**3**). With the addition of the TMO reagent both methyl groups and H^+ ions are introduced to the system. If the methyl group interacts with the carbamate the modification becomes trapped (**4**) however the production of H^+ ions causes dissociation of the formed carbamate and returns the amine group to its original charged state (**5**-**8**) (Scheme 4-8).



Scheme 4-8 Process of possible equilibria present with acetyl-lysine in solution with the addition of CO_2 and TMO.

The most likely hypothesis for the reaction being unsuccessful is therefore the introduction of H^+ ions by the hydrolysis of TMO causing the dissociation of the carbamate.

To improve the disruption of the experimental pH the next experiments were carried out using a pH stat. A pH stat is a machine able to measure the pH throughout the experiment and slowly add 1 M NaOH to buffer the pH around a set value. It was hypothesised this would cause a much smaller range of pH fluctuation during the experiment. The trapping reagent was also changed, TMO was exchanged for triethyloxonium tetrafluoroborate (TEO) which has a much longer half-life (117).

4-3.3 Amount of possible carbamate formation

The ε-amino group of lysine has a pKa value in the range 9.2-10.5 depending on the environment (179). Therefore the trapped carbamate yields at pH 7.4 are expected to be low. Using the Henderson-Hasselbalch equation (Figure 4-6) describing the relationship between a base and its conjugate acid allows the percentage of free base in solution to be calculated. Using the pH of the reaction 7.4 and the pKa of a lysine amine group as 10.5 (180) then using the equation below the ratio of freebase to conjugate acid is 0.0008. This means that under these conditions 99.9 % of the acetyl-lysine will be in the conjugate acid form.

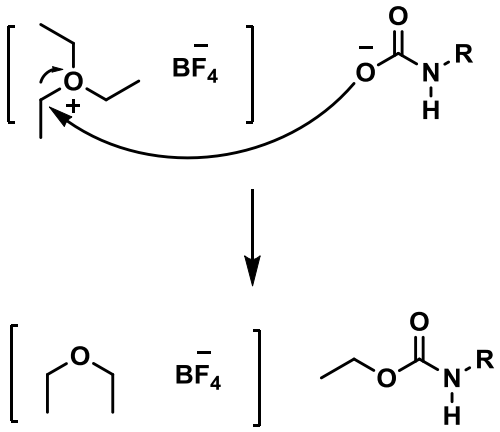
$$\boldsymbol{pH = pKa + log_{10}\frac{[B]}{[BH^{+}]}}$$

Figure 4-6 Henderson-Hasselbalch equation.

The use of this equation demonstrates a very low amount of lysine that would be capable of forming a carbamate. This factor suggests that the pKa of a lysine capable of forming a carbamate has been altered by the structure of the protein and the hydration space of its environment. This is true in the case of the *N*-terminal of Hb which is known to have a pKa of 6.6.

4-4 Trapping carbamates using TEO

TMO is a very reactive reagent in an aqueous environment, with a half-life of only 0.14 s (175). This means that the reagent is quickly hydrolysed during the reaction, decreasing the pH. TEO has a much longer experimental half-life of 7.4 min (117) making it a suitable alternative for investigating carbamate trapping under physiological conditions. The increased half-life of hydrolysis means that the production of H^+ ions is much slower and therefore easier to control during the reaction. The mechanism is very similar to TMO except an ethyl group instead of a methyl group is transferred (Scheme 4-9).

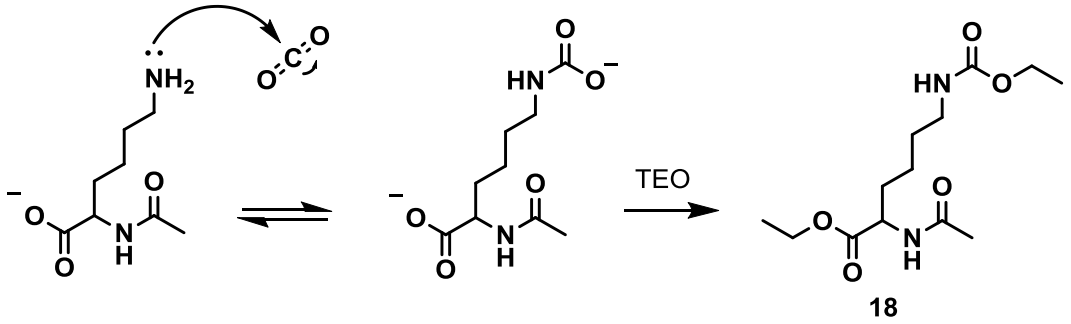


Scheme 4-9 Transfer of an ethyl group from TEO to a carbamate.

4-4.1 Acetyl-lysine trapping

It was demonstrated (Figure 4-3) via ^{13}C NMR that a carbamate is formed on acetyl lysine under the experimental conditions being used. An experiment was then performed to investigate whether the TEO reagent was able to trap the carbamate via ethyl transfer. As described previously, the trapping experiment was carried out with physiologically relevant levels of $NaHCO_3$ and pH 7.4 (181,182). The previous experiments described here used the manual addition of NaOH (117), however a pH stat was now introduced for more accurate pH maintenance (176).

As discussed previously, transfer of the ethyl group to the carboxyl group as well as the carbamate is expected. The hypothesised trapped product is shown in Scheme 4-10.



Scheme 4-10 The process of carbon dioxide binding to the ε-amino group of the lysine to form a carbamate and the predicted results of the addition of TEO on both the carbamate and the carboxyl group.

TEO was added with stirring in three incremental amounts to limit the disruption of the pH of the solution, while NaOH (1 M) was added via a pH stat. The pH stat monitors the pH over the course of the reaction and an example readout is given in Figure 4-7. The pH is kept stable over the course of the experiment with some fluctuations seen with reagent addition which are artefacts due to mixing.

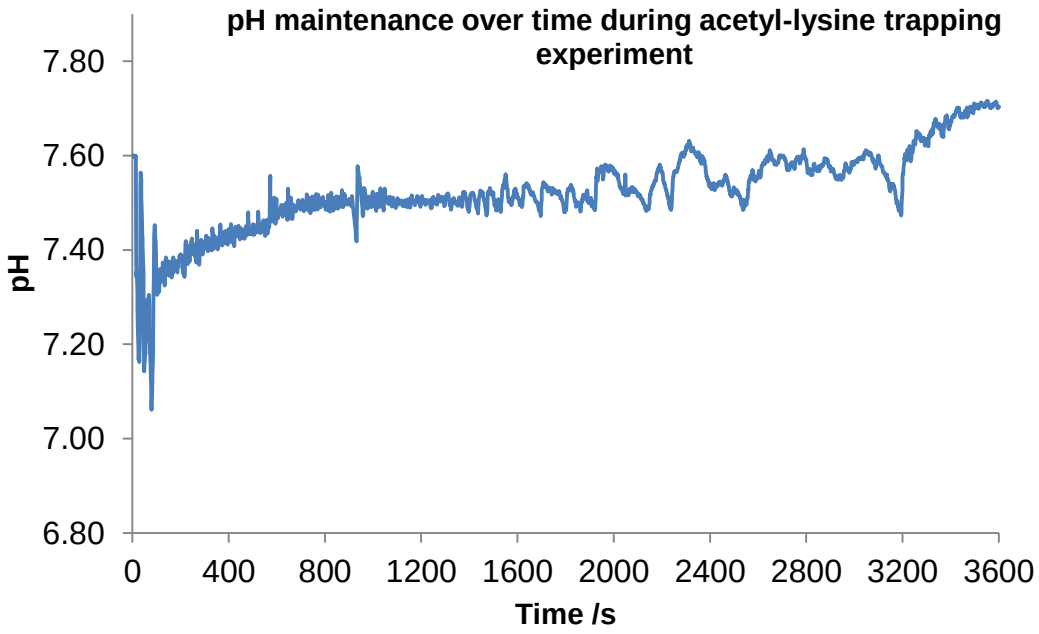


Figure 4-7 Graph produced by pH stat showing the maintenance of pH over an acetyl-lysine trapping experiment.

The results of a TEO acetyl-lysine trapping experiment were examined using ESI-MS (Figure 4-8). The trapped carbamate product was identified (**C**) as well as starting material with one transferred ethyl group on the carboxyl terminus (**A**) and starting material containing an ethyl group on both the carboxyl and amine groups (**B**). These peaks were assigned using the mass to charge ratio of the molecules. Figure 4-8 shows the total ion chromatogram (TIC) displaying the three peaks of products **A**, **B** and **C** as well as the m/z spectrum for the trapped carbamate product (**C**) both with and without the presence of a sodium ion.

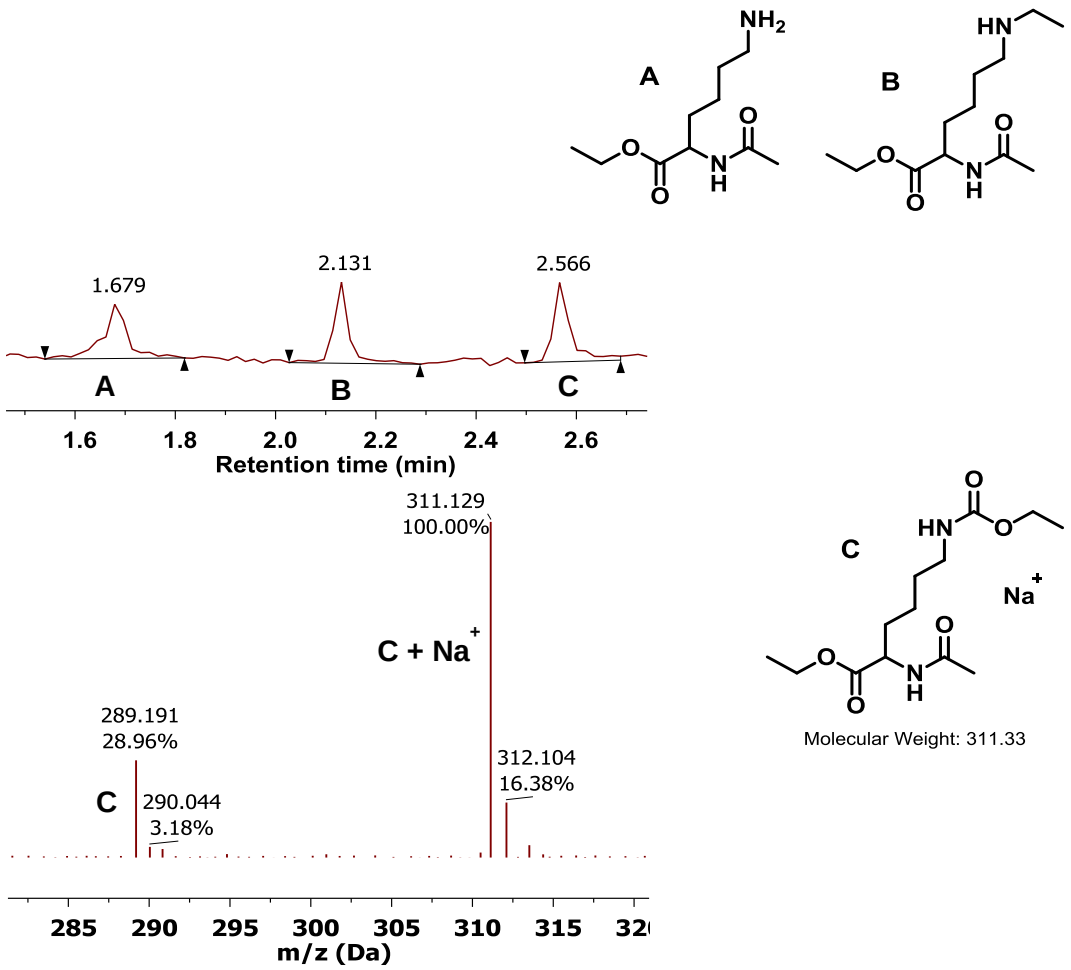
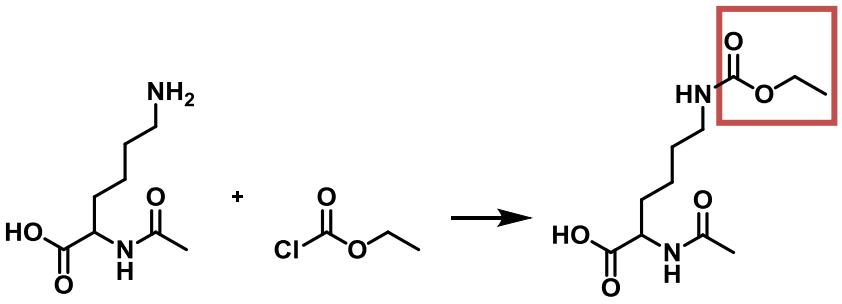


Figure 4-8 TIC and m/z trace from trapping of acetyl-lysine with TEO. Showing the products to be acetyl-lysine with ethylation of the carboxyl group (A), ethylation on both the carboxyl and amine groups (B) and the trapped carbamate on the ε-amine of acetyl-lysine (C). The mass of the trapped acetyl-lysine C – 289 and the 13C isotope 290 is seen in the m/z chromatogram.

This result demonstrates that under these experimental conditions it is possible to form a carbamate on an amine group under physiologically relevant conditions of pH and [CO_2] and trap it by ethyl transfer with TEO. This transfer of an ethyl group makes the modification robust enough for analysis by MS.

4-4.2 Synthesis of synthetic carbamylated acetyl-lysine

For further confirmation that the product (**C**) identified corresponds to a trapped carbamate, product **C** was purified from the side products by extraction into ether (2-2.2.1) and compared by NMR spectroscopy to a chemically synthesised product (Scheme 4-11).



Scheme 4-11 Synthesis of chemically modified acetyl-lysine derivative containing a group chemically equivalent to a trapped carbamate.

This synthesised acetyl-lysine derivative contains a group chemically equivalent to a trapped carbamate (red box). This material was compared to the purified trapped material (**C**) by ^{1}H NMR spectroscopy (Figure 4-9).

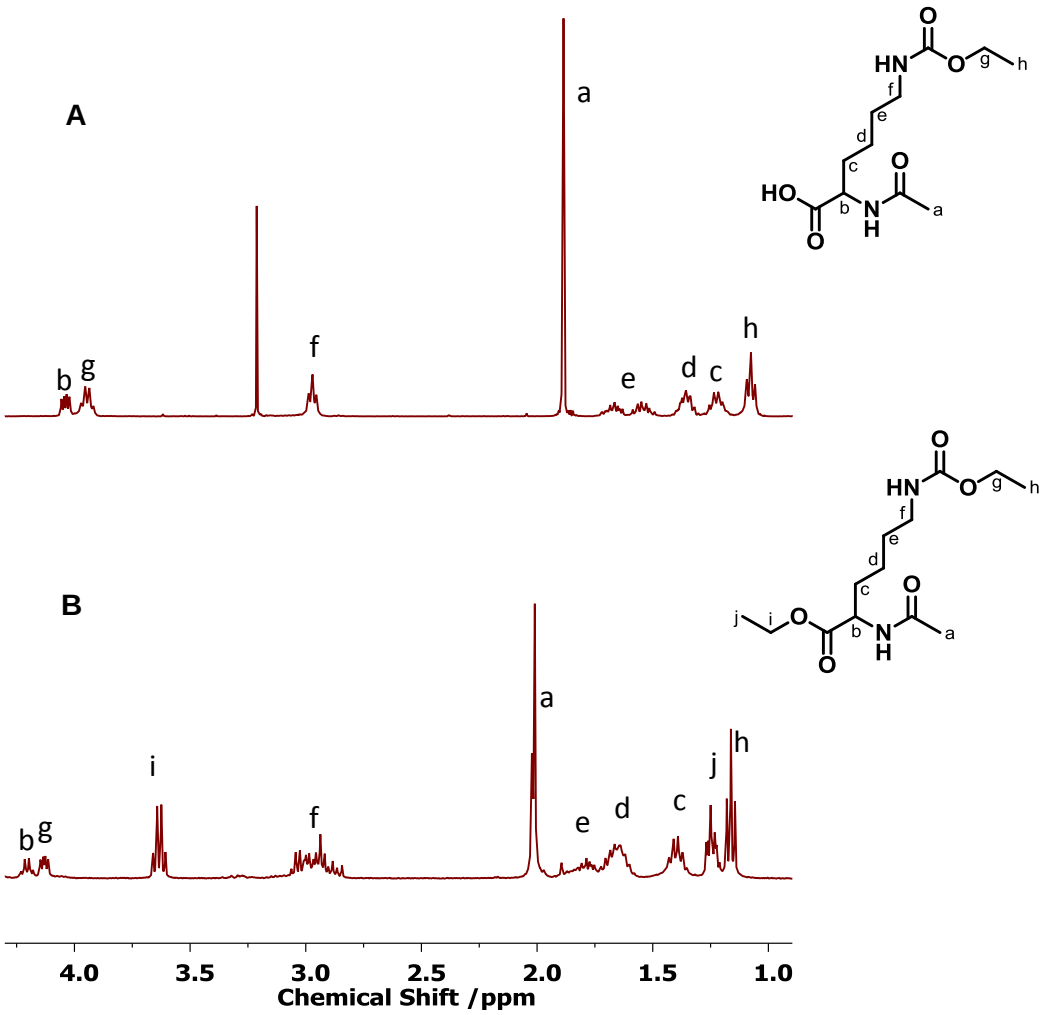


Figure 4-9 ^{1}H NMR comparing signals between the chemically carboxyl ethylated acetyl-lysine (A) and the material formed during the trapping reaction (B).

The product made by the trapping reaction includes an additional ethyl signal on the carboxyl group accounting for signals i and j (**B**).

Though the separation of the three products (**A**, **B** and **C** Figure 4-8) can be seen on the TIC from the mass spectrometry trace the purification of **C** (the carbamate trapped desired product) proved difficult by silica chromatography due to the similar affinity of the three compounds. This introduced some contamination signals around signal f in spectrum **B** (Figure 4-9) corresponding to side products **B** and **C** in Figure 4-7. It was not possible to completely remove this contaminant.

Despite these two differences the spectra match very closely and confirm that product **C** from the trapping reaction is acetyl-lysine with a trapped carbamate.

4-4.3 ^{13}C Confirmation of carbamylated acetyl-lysine

It is assumed in the carbamate trapping reaction that the $NaHCO_3$ is dissociating within the aqueous system to provide CO_2 for carbamate formation. To confirm that this is occurring a labelled sodium bicarbonate source ($NaH^{13}CO$) was used as an orthogonal approach. This method has been used previously to investigate carbamylation in β-lactamases and adult human Hb in NMR analysis (109,183).

The m/z of this experiment was then compared to the ^{12}C experiment to look for a mass shift (Figure 4-10).

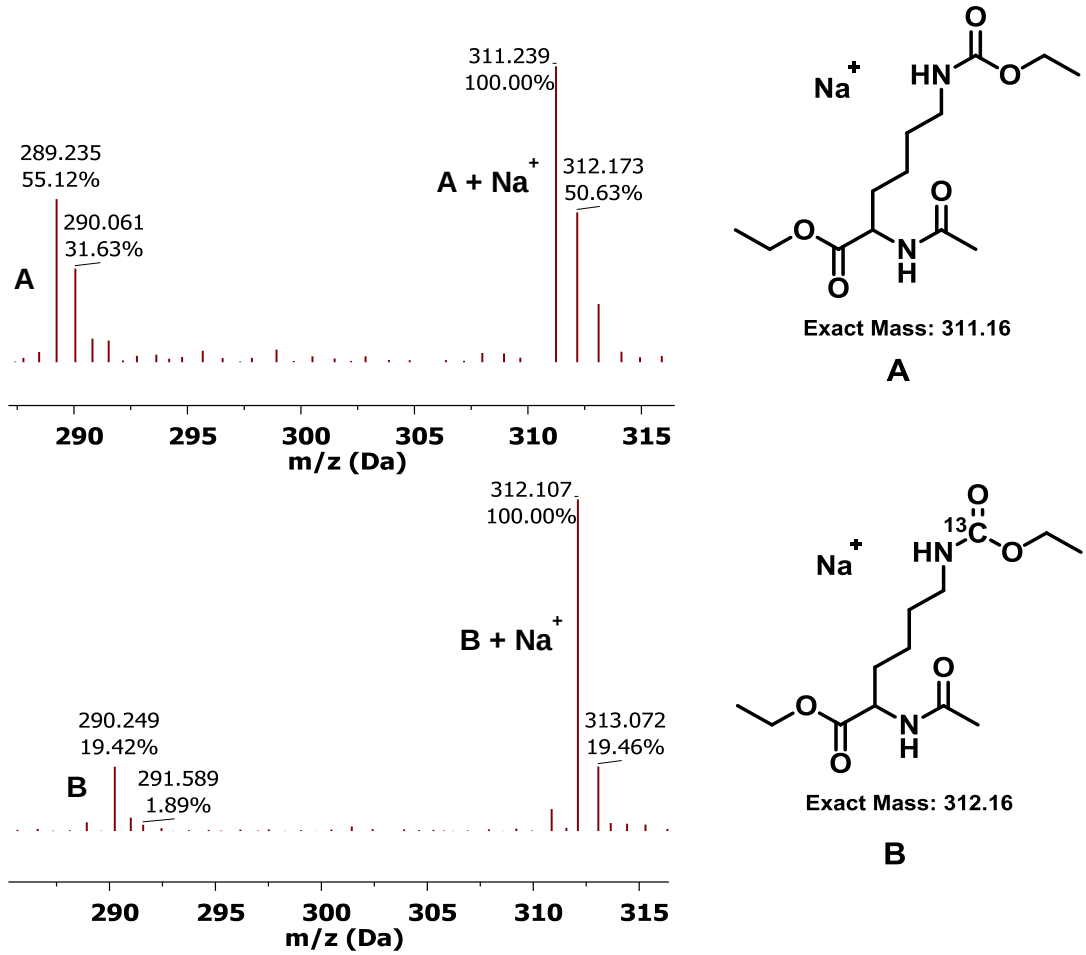


Figure 4-10 MS trace showing the increase of one mass unit from ^{12}C (A) with the use of ^{13}C labelled sodium bicarbonate (B).

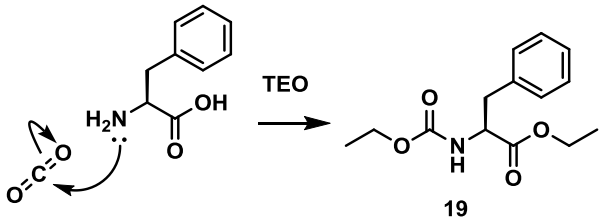
The expected 1 Dalton (Da) mass shift increase seen between ^{12}C trapping (product **A**) and ^{13}C trapping (product **B**) confirmed that it is the CO_2 being formed in solution under the experimental conditions that is causing carbamate formation.

4-4.4 Phenylalanine trapping

Due to the difficulty removing the acetyl-lysine side products the yield of the trapping reaction could not be accurately confirmed. The amount of substance in a reaction cannot be compared by MS because of effects such as ion suppression (184).

Without the separation of the pure trapped product, another method to discover the yield of carbamate trapping is to use a UV active group so that the amount can be

analysed from the MS UV trace. To introduce a UV active group the experiment was repeated on the amino acid phenylalanine (Scheme 4-12).



Scheme 4-12 Formation and trapping of a carbamate on the *N*-terminus of phenylalanine.

Phenylalanine does not have an amine group on its side chain but carbamate formation is also possible on an *N*-terminus α-amine group, such as occurs on Hb (138). However this is not a direct comparison with acetyl-lysine because an *N*-terminus amine group has a lower pKa of around 9 (180) so will more readily form a carbamate under the same pH conditions. Using the Henderson-Hassellbach equation the ratio of freebase to conjugate acid is 0.025. This means that under these conditions 97.6 % of the α-amine group will be in the conjugate acid form.

The experiment was carried out under the same conditions of $NaHCO_{3,}$ pH and TEO as the acetyl-lysine trapping and the results were analysed by ESI-MS (Figure 4-11).

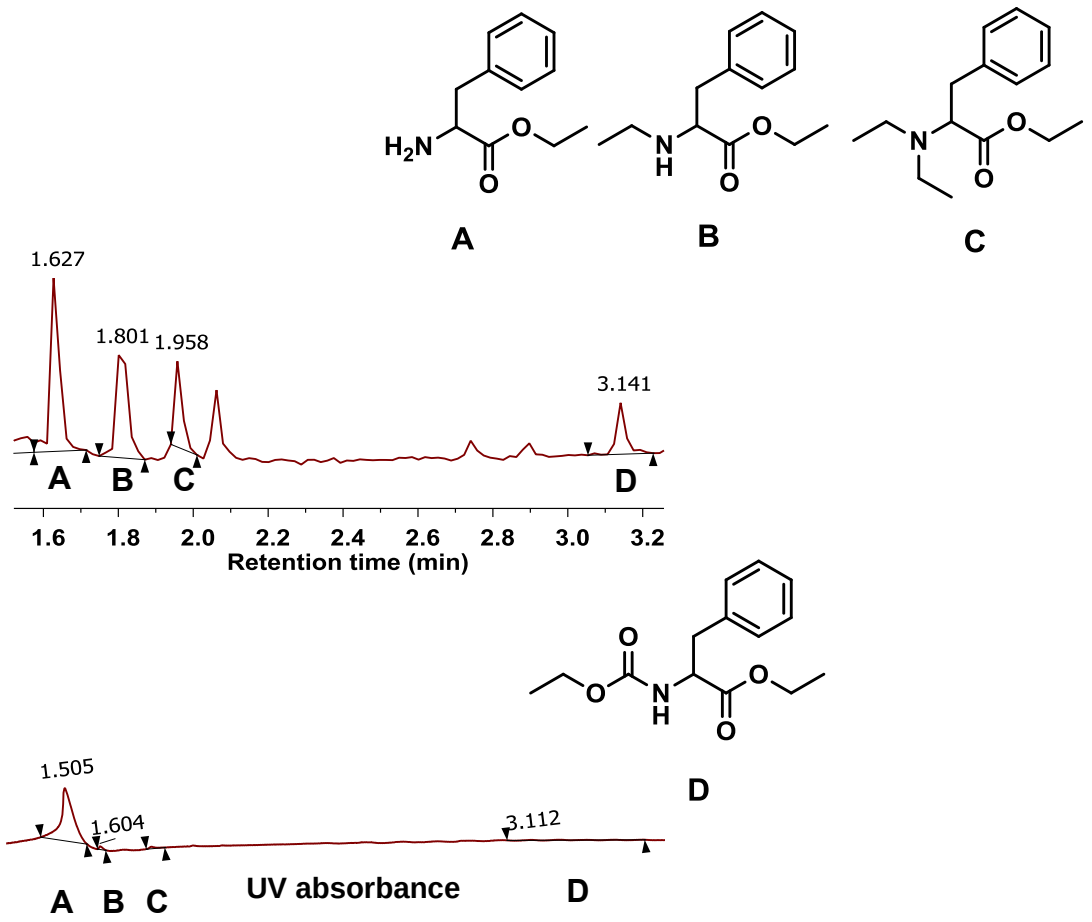


Figure 4-11 TIC and UV traces of trapped phenylalanine reaction at pH 7.4 showing the carbamate trapped product D and ethylated side products A, B and C.

A carbamate was successfully trapped on the *N*-terminus α-amine of phenylalanine ((**D**) Figure 4-11). There was also the starting material with ethyl transfer to the carboxyl group (**A**) and double transfer to both the carboxyl and the amine group once (**B**) and twice to the amine group (**C**). The yield was calculated by comparing the area of the product **D** to side products **A**-**C** on the UV trace (Figure 4-11).

Phenylalanine experiments were carried out at pH 7.4 and 8.5 to investigate the effect on carbamate formation of the pH of the reaction (pH 8.5 not shown). Both experiments contained the same product and three side products. The yield was compared using the MS UV system with a minimum threshold of 0.5 %. Though the 'peaks' seen by the software (Figure 4-11 UV absorbance spectra) are small enough to offer a large possibility of variance the difference between the peak observed at

pH 8.5 and pH 7.4 was large. This goes to support the hypothesis that the reaction is pH limited due to the charge of the amine group.

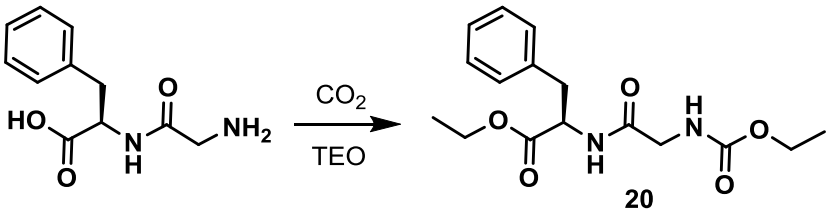
This demonstrates the importance of the pH of the solution and therefore the charge state of the amine group in the formation of a carbamate. It is known that the rate-limiting step in the formation of a carbamate is the binding of CO_2 to the amine (78) which is encouraged by a neutral charge state of the amine group. As normal physiological environments are at pH 7.4 it is hypothesised here that carbamate sites are contained within privileged protein environments which encourage neutral amine formation.

4-5 Development of the trapping method on dipeptide and tetra peptide systems

After successfully trapping a carbamate on a single amino acid system, the next step was to examine carbamate trapping on a dipeptide. Two different dipeptides were used both containing phenylalanine; the second amino acid was glycine as a control molecule without a side chain amine and lysine. This was to compare carbamate formation on a lysine side chain versus an *N*-terminus amino group.

4-5.1 Glycine-Phenylalanine (GLY-PHE)

The first investigation of a dipeptide system used the dipeptide GLY-PHE. The same amount of $NaHCO_3$ and TEO were used and the expected reaction is shown in Scheme 4-13.



Scheme 4-13 GLY-PHE dipeptide reaction with CO_2 and TEO and its product after trapping with the formation of a carbamate.

These results were analysed using ESI-MS (Figure 4-12).

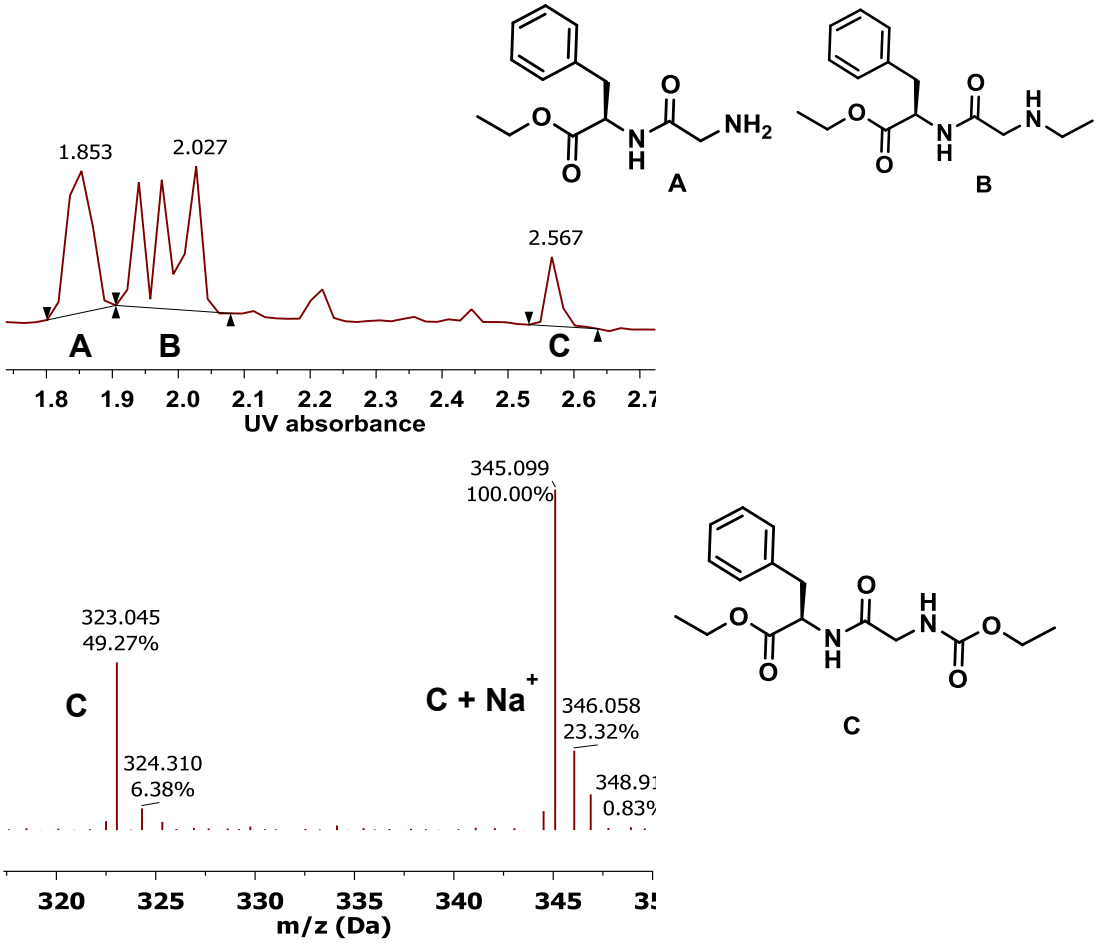


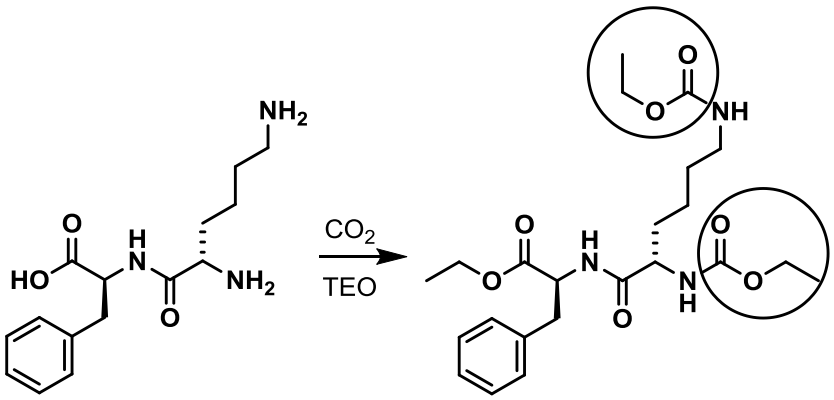
Figure 4-12 ESI-MS showing the trapped carbamate product of GLY-PHE (C) as well as the singly ethylated starting material (A) and doubly ethylated starting material (B).

A carbamate was successfully trapped on the N-terminus of the dipeptide (**C**) as expected the reaction also created the side products of single and doubly ethylated GLY-PHE (Figure 4-12, **A** and **B** respectively). This successful result has confirmed that the method has progressed to be able to form and trap a carbamate on a dipeptide molecule for analysis by MS.

4-5.2 Lysine- Phenylalanine (LYS-PHE)

Next, an investigation into carbamate formation on an α-amine versus an ε-amine group was carried out using the dipeptide LYS-PHE. Both have been seen on proteins in literature (the Hb carbamate is on an N-terminus α-amine group (90) and the RuBisCO carbamate ε-amine K201 (78)).

This experiment was carried out using the same developed conditions and analysed using ESI-MS. The hypothesised reaction can be seen in Scheme 4-14 with both possible carbamate locations highlighted.



Scheme 4-14 LYS-PHE dipeptide and its possible carbamate formation sites (circled) with trapping.

This experiment was analysed using ESI-MS (Figure 4-13).

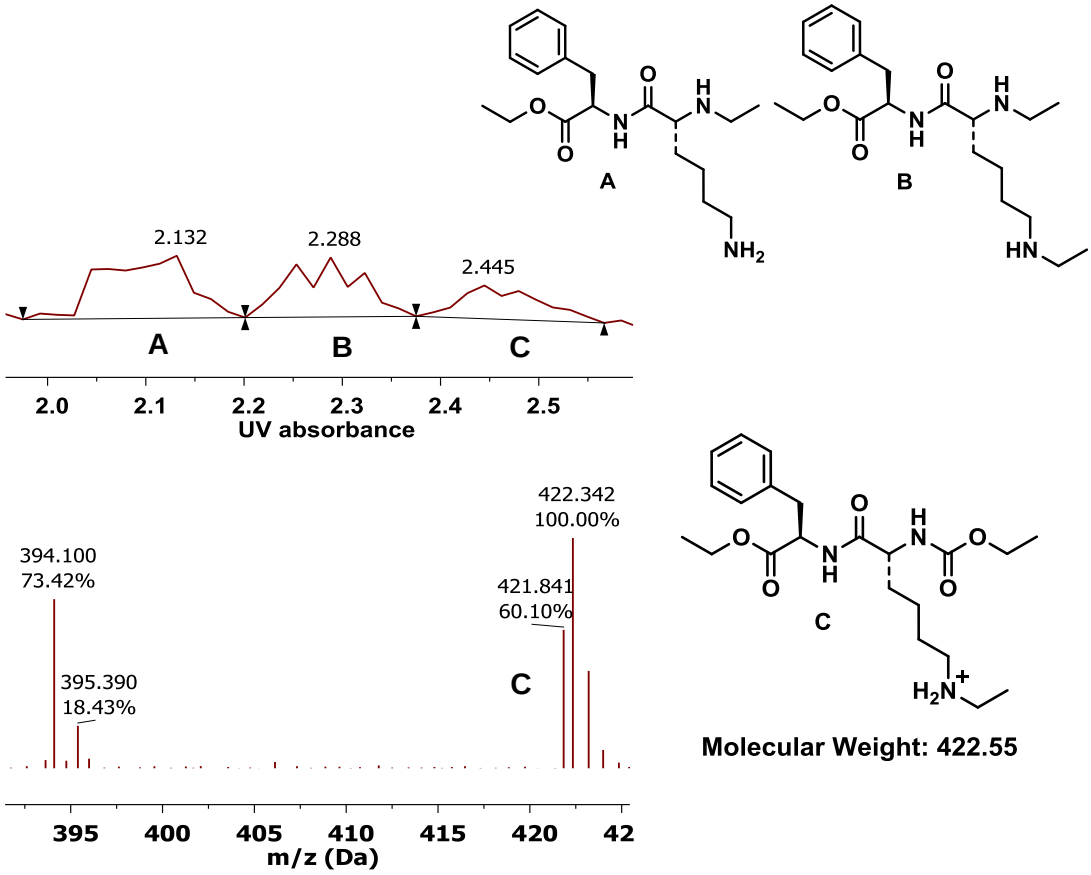
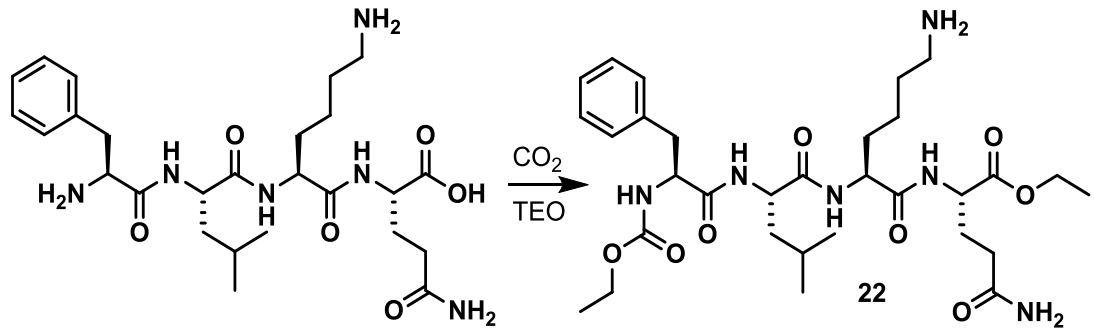


Figure 4-13 LYS-PHE MS spectrum showing the side products with additional ethylation (A and B) and the trapped carbamate product with one and two ethyl additions (C).

The results of this experiment showed a successfully trapped carbamate product peak of 422 m/z (Figure 4-13, **C**). However, no doubly carbamylated product was found, most likely due to steric hindrance. The MS could not be used to say which location the carbamate formed at but due to the pKa values, it is most likely to be the *N*-terminus amine. In this case the side products were doubly and triply ethylated starting material (Figure 4-13, **A** and **B** respectively).

4-5.3 Phenylalanine-leucine-lysine-glutamine (FLKQ) Tetrapeptide

After development of the method on a dipeptide system, a tetrapeptide (FLKQ) was investigated. The process of reaction of FLKQ with CO_2 and TEO with carbamate formation expected on the N-terminus is shown in Scheme 4-15.



Scheme 4-15 FLKQ tetrapeptide reaction with CO_2 and TEO.

This reaction was successful and produced a trapped carbamate on the N-terminus of the tetrapeptide (Figure 4-14 **A**) and a trapped carbamate molecule with an additional ethylation on the lysine side chain (Figure 4-14 **B**).

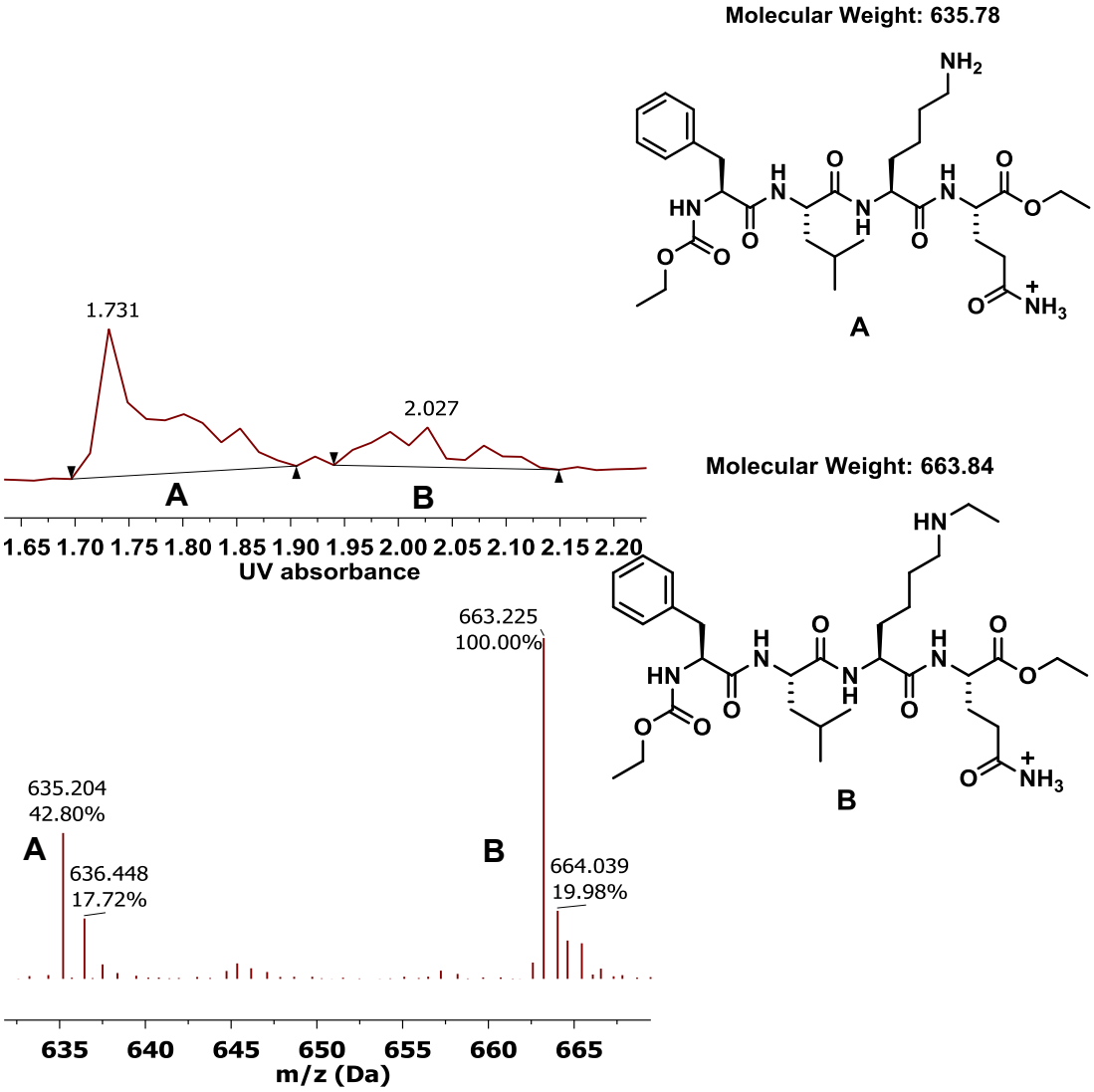


Figure 4-14 FLKQ tetrapeptide MS results showing a trapped carbamate on the N-terminus with an ethylation on the C-terminus (A) and on the lysine group (B).

Once the method had reached a stage of repeatedly trapping a carbamate on a small molecule, the next step was to develop the method on a whole protein.

4-6 Haemoglobin trapping

The overall aim of this investigation is to develop a method for screening carbamates on a proteome. Therefore, it is important to develop this method to be able to trap carbamates within a whole protein.

Hb was chosen as the first full protein for development of the method because it is one of the few proteins already confirmed to form a carbamate (79). The formation of the carbamate is known to be on the N-terminus residue of the β–chain and therefore an easily accessible site.

First the formation of a carbamate was confirmed on Hb using ^{13}C NMR spectroscopy and compared to literature values (81). The formation of a carbamate on Hb was confirmed in the same way as on acetyl-lysine, by the appearance of a signal at 164 ppm (**A**) in ^{13}C NMR spectroscopy (Figure 4-15).

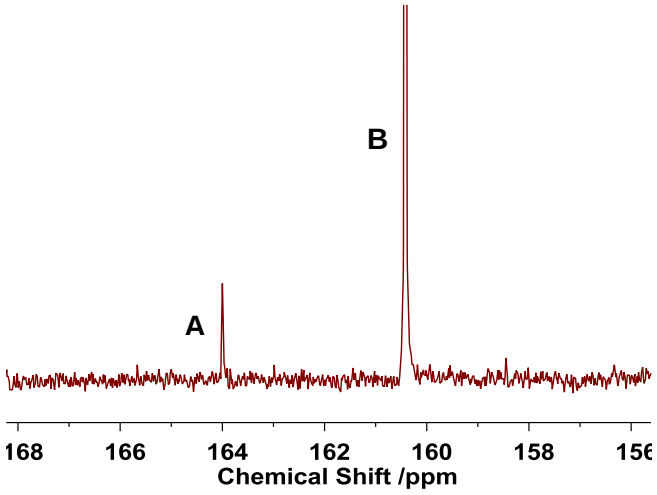
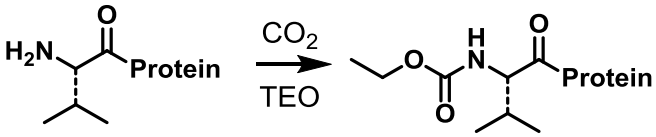


Figure 4-15 ^{13}C NMR spectrum showing a bicarbonate in solution peak B and a carbamate peak A.

The site of carbamate formation on Hb is on the α-amine of the β chain N-terminus residue valine (138). The carbamate formation and trapping on this residue can be seen in Scheme 4-16.



Scheme 4-16 Trapping of a carbamate formed on the N-terminus valine residue of Hb.

The Hb experiment was carried out using purchased human Hb (Sigma Aldrich H7379) and the same reaction conditions previously developed. Applying the method to a whole protein produced several new challenges.

4-6.1 Proteomics

Proteomics is the study of the whole proteome, all the proteins expressed within a cell including their modifications (185). The study of protein PTMs requires the use of MS to analyse the modification (186). MS used to only be applicable to small molecules but the introduction of new 'soft' ionisation techniques made protein MS possible (187). There are two main types of 'soft' ionisation used for protein MS, electrospray ionisation (ESI) and matrix assisted laser desorption ionisation (MALDI) (188). Both of these MS techniques were used in this investigative work. MALDI MS uses a solid phase matrix to assist in ionisation, in this work the matrix used was alpha-cyano-4-hydroxycinnamic acid (α-CHCA) (189). The matrix is a molecule that is designed to absorb the laser energy and assist the sample in ionisation. ESI works by forcing the liquid sample out of a needle tip causing it to disperse into a fine spray of charged droplets (190). ESI is easily coupled to become tandem MS (MS-MS) which provides additional peptide fragmentation information necessary for sequencing.

Though these new soft ionisation techniques allow the visualisation of most PTMs but due to the labile nature of carbamylation it is still difficult. The trapping of the carbamate has made this modification robust enough to view by MS.

Proteins must be digested for MS because peptides are more easily soluble and ionise better than whole proteins. The digest was carried out using the protease trypsin which cleaves on the C-terminus side of lysine and arginine residues (191). The full description of how the digestion was developed is covered in chapter 5.

4-6.2 MALDI

During the protein trapping reaction TEO also transfers ethyl groups to other amino acids in Hb. This becomes more of a challenge when investigating a complex protein mixture and is discussed more thoroughly in chapter 5.

MALDI data comes in the form of a list of mass peaks to be analysed. As the location of the Hb carbamate is known to be on the β chain N-terminus peptide the expected mass of interest can be calculated.

The sequence of the β chain N-terminus peptide is VHLTPEEK and the mass of this peptide is 952.51 Da. Ethyl transfer was expected on both E amino acids (each ethylation adds 28.03 Da) which would give the mass 1008.57 Da. This mass was therefore the starting point for the search of the modified peptide in the MALDI data. The addition of a trapped carbamate (CO_2 and an ethyl group) adds 72.02 Da. The addition of an ethyl group is a variable modification so a series of products is possible (Table 4-1).

Table 4-1 Showing the mass peaks from the MS in Figure 4-16 and their corresponding modifications.

Peak m/z	Non-carbamate series	Trapped carbamate series
1008.57	N-terminus + 2 ethylations	-
1023.61	-	-
1036.60	N-terminus + 3 ethylations	-
1064.64	N-terminus + 4 ethylations	-
1080.59	-	N-terminus + carbamate + 2 ethylations
1092.67	N-terminus + 5 ethylations	
1108.62	-	N-terminus + carbamate + 3 ethylations
1120.70	N-terminus + 6 ethylations	
1136.66	-	N-terminus + carbamate + 4 ethylations

The mass of the N-terminus peptide with a trapped carbamate and two ethylations is 1080.59, which can be seen in the MALDI spectrum (Figure 4-16). The MALDI data shows an ethylation addition series for the N-terminus peptide (red peaks) and for the trapped carbamate peptide series (blue peaks).

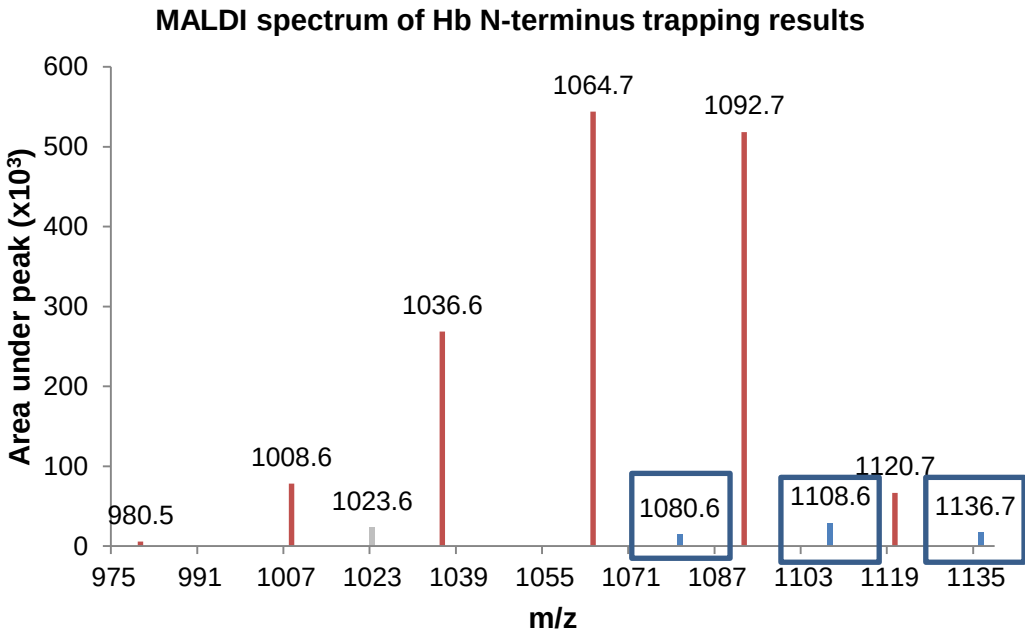


Figure 4-16 Hb MALDI spectrum, showing both an ethylation series (red peaks) and a trapped carbamate ethylation series for the N-terminus peptide (blue peaks). The trapped carbamate series is highlighted in blue boxes and increases by the mass of an ethylation (28).

Figure 4-16 shows the ethylation series (red peaks) as having a large relative intensity compared with the ethylation series also containing the trapped carbamate (blue peaks). However as demonstrated by the phenylalanine experiment the intensity level of peaks does not always correspond to the experimental values.

Originally in literature (173) the amino acids thought to be ethylated by TEO were only glutamate and aspartate but looking at the length of the ethylation series in these results (showing at least four additional ethyl groups) there are more ethylated amino acids than previously suggested.

The peptide suspected of being the N-terminus residue mass plus a carbamate + 2 ethyl groups (1080.59 Da) was fragmented for sequence confirmation. Fragmentation is caused by increased energy inside the MS machine which creates ionised fragments from a digested peptide, these are then used to gain sequence information about the peptide (Figure 4-17).

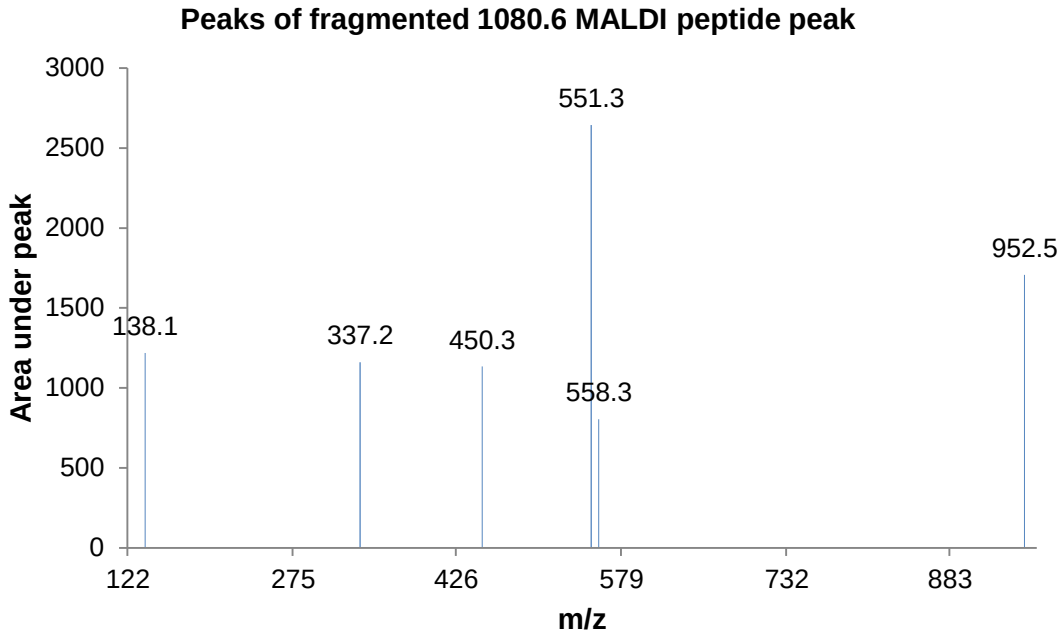


Figure 4-17 MSMS fragmentation peaks of the N-terminus peptide peak 1080.6.

These fragmentation peaks were then compared to possible mass fragmentation of the peptide VHLTPEEK + carbamate + two ethylations. The fragments that correspond to the peaks in Figure 4-17 are detailed in Table 4-2.

Table 4-2 Comparing the m/z fragments from 1080.6 fragmentation to the N-terminus peptide sequence.

Peak m/z	Corresponding amino acid sequence
138.1	H
337.2	VH + carbamate + ethylation
450.3	VHL + carbamate + ethylation
551.3	VHLT + carbamate + ethylation
558.3	PEEK + 2 ethylations
952.5	VHLTPEEK

This fragmentation confirms the presence of a carbamate on the N-terminus peptide.

Due to the wide range of additional ethylations the MALDI spectra contains many peaks. This creates difficulty when manually searching the data for sites of carbamylation and was only possible for Hb due to previous knowledge of the location of the carbamate. ESI was therefore used as an alternative technique. ESI can easily be coupled to an MS-MS system so that peptide fragmentation can be collected inline. MALDI also cannot be used to confirm a modification in the first instance because, unlike ESI, it does not collect MS-MS data which is needed to provide sequence information.

4-6.3 ESI

ESI acquisition takes longer due to collecting MS-MS data. However, this provides much better analysis and confirmation of modifications. Further details of the MS techniques explored in this work are detailed in chapter 5. ESI MS contains MS-MS fragmentation caused by increased energy; this allows the identification of the amino acid sequence and in this case the additional modifications being carried. Previous literature surrounding carbamates have tried to analyse via MS but this increased energy has removed the carbamate, with this new trapping method the modification can be retained (78).

The use of ESI meant that the data collected needed processing software. There are many MS online software available but they are all limited to a maximum of two variable modifications so were unusable for this work. A collaboration with Newcastle University introduced the software GPM X!Tandem (Tandem) which is able to handle multiple modifications. ESI-MS produces MS-MS data which Tandem can then compare to a specified genome to search for the detailed modifications. A score is provided along with the peptide and modification assignment to provide information about the likelihood of a false positive. In this case a false positive would correspond to a fragment ion being assigned to the wrong peptide.

The search was carried out including the carbamate modification (within Tandem referred to as carboxyethyl) on lysine and protein N-terminals and the ethylation modification (within Tandem referred to as dimethyl) occurring on glutamate and aspartate residues. The ethylation can also occur on other amino acids with free

electrons such as lysine and arginine but a larger number of modifications increases the chances of a false positive so these were omitted in the original search.

The software successfully found a mass corresponding to the N-terminus peptide with a carbamate and one ethyl group (1052.56) from the ESI data (Figure 4-18). The score for this hit was high and there were no other carbamates found on the protein.

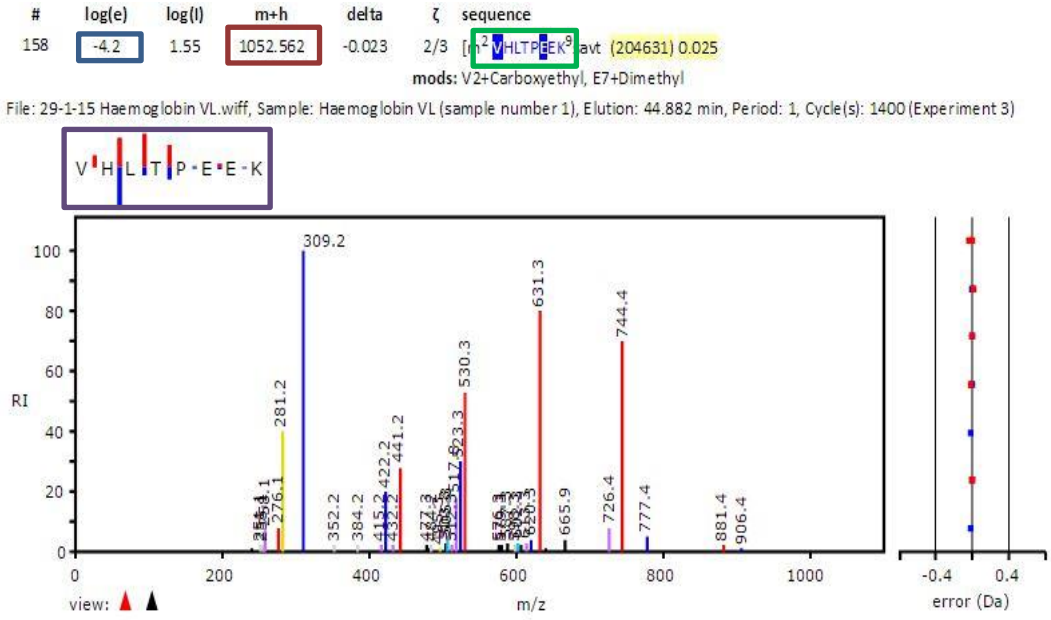


Figure 4-18 Tandem image containing the N-terminus peptide with a trapped carbamate on the valine residue (green box) a good score (blue box) and the mass for the peptide (red box).

The Tandem software allows the user to search through the data by looking for a specific PTM, in this case a trapped carbamate (labelled carboxyethyl by the software). The software then displays the peptide sequence containing the modifications (green box), the mass of the peptide (red box) and the confidence score (blue box) to demonstrate the assignment. The fragments found in the data that match the peptide (purple box) and shown as blue peaks for b ions and red peaks for y ions which correlate with the peaks shown in the spectra.

To confirm that the formation and trapping of the carbamate is due to the experimental procedure, an isotope experiment was completed using $NaH^{13}CO_3$. $NaH^{13}CO_3$ has previously been used to investigate Hb carbamate binding (136) and would confirm the carbamate formation by the appearance of the same peak but

with an additional 1 Da mass shift. The experiment was successful and the expected increase of one mass shift from 1024.10 VHLTPEEK + CO_2 to 1025.11 VHLTPEEK + $^{13}CO_2$ can be seen in Figure 4-19.

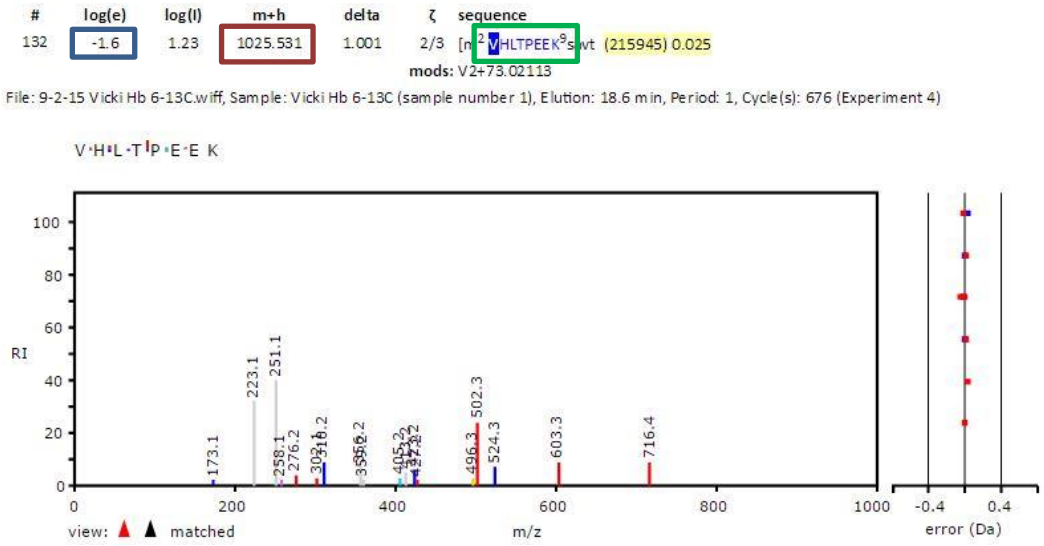


Figure 4-19 Tandem image showing the N-terminus peptide with a ^{13}C trapped carbamate being 1 dalton heavier than the ^{12}C sample mass of 1024.5.

A control was also carried out to investigate the hypothesis that carbamates are able to form because of the privileged environment of the folded protein. To control against this a trapping experiment was undertaken after the Hb protein had been denature in 4 % SDS to unfold the protein before trapping. The majority of the SDS was removed from the reaction before trapping. No trapped carbamate was observed (Figure 4-20) within this sample, indicating that the structure of the protein is necessary for carbamate formation.

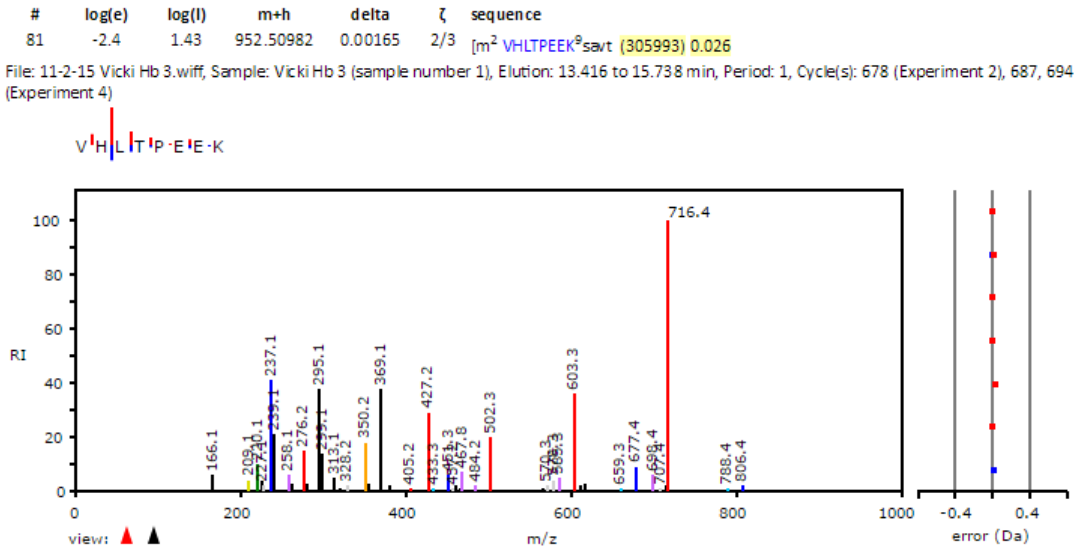


Figure 4-20 Tandem image for a control experiment showing the N-terminus peptide without a carbamate, mass 952.5 and a high confidence score of -2.4.

These results demonstrate that the method has now been developed to a stage where it can be used to trap a carbamate, formed under these conditions, for downstream MS analysis on a whole protein.

4-7 Conclusion

Previously there had been little work done on the investigation of the post-translational modification carbamylation. The main reason behind this area being so understudied is the labile nature of the modification. The common method of PTM analysis uses MS which, due to its high energy ionisation and acidic conditions is unsuitable for carbamylation identification. This has meant that all known carbamates have been confirmed using non-native conditions which could cause artefacts. This work presents the development of a method to overcome these restrictions. This method can be used on proteins under native aqueous conditions to convert a carbamate into a modification robust enough for in-depth analysis.

The work in this chapter has described the development of a carbamate trapping method for use under physiological conditions. This work builds on previously established alkylation chemistry but under aqueous conditions. The work began by trapping a carbamate on a single amino acid system using acetyl-lysine and progressed through two dipeptides and a tetrapeptide. Once a stable trapping method had been confirmed the work progressed on to a whole protein. Hb was the protein chosen for this due to the repeated confirmation of carbamate formation within literature (136,140,192). The results were analysed using MALDI and ESI MS and were confirmed with ^{13}C isotope labelling.

This method can now be applied to any protein and can be used in all systems to search for novel CO_2 target proteins, thus opening up a pathway to further understand the complex interactions of CO_2 within a cellular environment. Carbamate formation on Hb and RuBisCO are important for function of the protein. Therefore it is hypothesised that this modification could also be important in the function of other proteins.

4-8 Future work

This chapter has described the successful development of a method that can trap a carbamate under physiological conditions on a protein.

The overall aim of this investigation was always to search for unknown carbamates. This developed method can now be used to carry out this aim. To investigate this the trapping method was used to screen an *Arabidopsis* soluble leaf lysate. This work and the development of the digestion and MS necessary for analysis is detailed in chapter 5.

Chapter 5: A proteomic screen for new protein carbamates

5-1 Overview

The overall aim of this project was to develop a system capable of searching for unidentified carbamylation sites on proteins under physiological conditions. This work began by developing a method to trap a known carbamate on a protein and is successfully described for Hb in chapter 4. Carbamates cannot be analysed via MS due to their labile nature. Development of a trapping method provides the ability to convert a carbamate into a robust modification amenable to analysis. This chapter describes the development of a method to trap unknown carbamates on a plant leaf lysate, with the aim of creating methods able to screen any proteome.

Arabidopsis thaliana (*Arabidopsis*) has been used as a model plant system since the 1980's (193). This is because it has a small sequenced genome and relatively short life cycle. The leaf lysate was chosen as the site for screening as this is the location of CO_2 interface through photosynthesis and therefore likely to contain proteins that are important for CO_2 interactions (194).

The leaf lysate is a far more complex sample than trapping on a single purified protein and therefore required methods of fractionation to simplify sample complexity before analysis. Several methods of fractionation were explored in this chapter and it was decided that fractionation after trapping was a suitable way to keep closest to physiological conditions. Analysis methods were developed to be able to search a complex range of modifications and criteria were developed to reduce the chances of false positives.

Newly discovered protein carbamates were discovered using two different analysis software packages. These carbamates were all on lysine residues at previously unknown sites and were identified by fragmentation patterns from ESI-MS data.

5-2 Previous work on complex protein systems

5-2.1 Previous proteome investigations

The predominant method used in proteomics screening is known as bottom-up proteomics (195). Bottom-up proteomics is the process of identifying proteins through tandem mass spectrometry (MS-MS) by digesting a protein sample to form peptides (196). In this work the soluble proteins from an *Arabidopsis* leaf were extracted and digested into peptides. These peptides were then fractionated and ionised using MS. These ionised peptides were then further fragmented with increased energy (MS-MS) (Figure 5-1). These new fragmented ions formed peaks corresponding to a mass to charge ratio (m/z) and were compared to a theoretical peptide spectrum from the selected database (197) which allowed protein sequencing and identification.

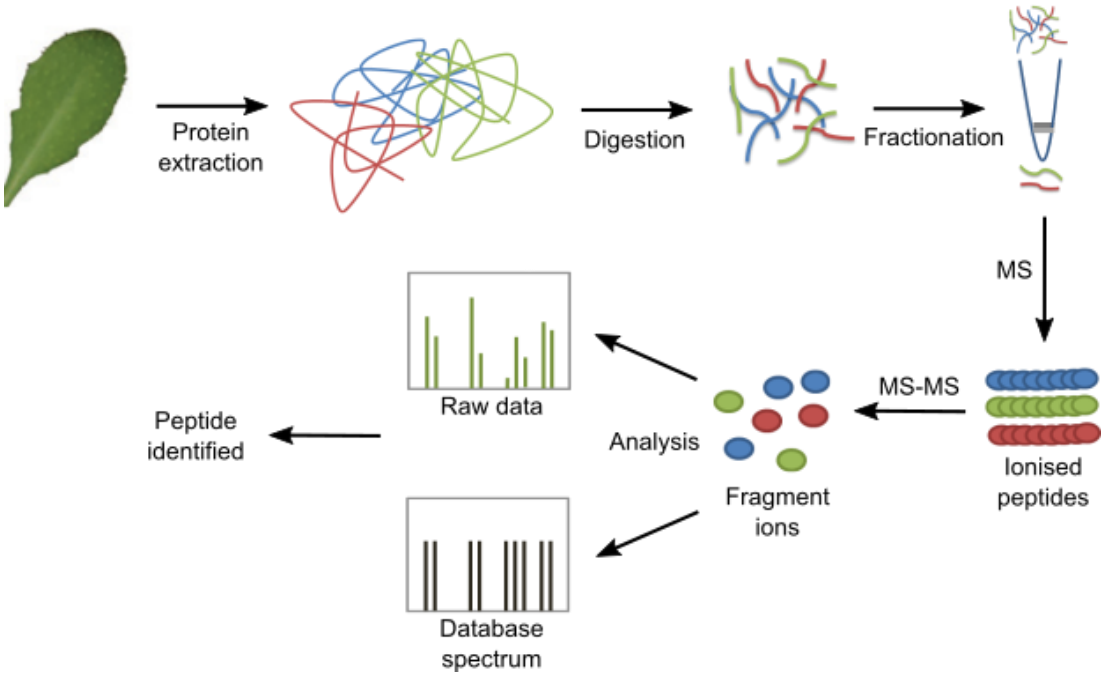


Figure 5-1 Steps involved in bottom-up proteomics for peptide identification from protein extraction.

Peptides are preferred over proteins for MS due to the higher sensitivity of MS for smaller molecules and the better fragmentation of peptides compared to proteins

(198). There are several proteases that can digest proteins, the most commonly used is trypsin because of its stability and low error rate (199). Trypsin works by cleaving on the carboxyl terminal of lysine and arginine residues (200) to form peptides.

There are several challenges associated with working on a large-scale proteome study. The largest of which is swamping of the sample by highly abundant proteins, which can cause rare proteins to be missed (201). In most studies the sample complexity is reduced by either fractionation or subcellular compartment enrichment (202). Recent studies investigating yeast proteomes were able to identify 3639 (203) and 3977 (204) proteins, roughly 2/3 of the theoretical yeast proteome, using LC-MS-MS triplicate runs when including fractionation before MS. The *Arabidopsis* genome encodes roughly 25000 proteins (205) so a method to reduce the sample complexity is essential.

Previous investigations of plant proteomes have focused on one cellular compartment; this is normally achieved by fractionating the lysate by density gradient centrifugation. This has been carried out in studies into both the chloroplast and mitochondrial proteomes of *Arabidopsis* where both studies begin by purification of their cellular compartments by Percoll density centrifugation (206,207). Next, sample proteins were separated into soluble and insoluble fractions and then additionally fractionated by either 2-DE (206) or by liquid chromatography (207). These studies identified 690 and 170 proteins respectively.

In most cases of proteome analysis the fractionation of the sample is targeted to a subcellular location in this way because the protein or PTM location is known and therefore the complexity can be easily reduced. In this work, it is hypothesised that carbamates are formed under any cellular environment and therefore can occur in all cell compartments so this method of directed fractioning was not used. Other methods of fractionation were explored here and are described in 5-3.2.

5-2.2 Previous work investigating PTMs

PTMs are changes to a protein which can alter the activity or localisation of that protein (208). The investigation of PTMs on proteins is paramount to understanding their biological function.

Previous screens have been carried out for both acetylation and phosphorylation within plants. Lysine acetylation was previously thought to only occur on histones but has recently been identified on 57 proteins within *Arabidopsis* (209). Several large scale studies of phosphorylation have been carried out in plants (202). A study of phosphorylation in an *Arabidopsis* whole cell lysate identified 1346 proteins containing modifications. Searching for both of these modifications can be improved by enrichment of the sample based on the modification. The acetylation study used immunocapture with anti-acetyl antibodies and the phosphorylation study used several enrichment methods including Fe-IMAC to remove unmodified peptides (210). Using enrichment allows the screening of a much larger sample as the sites of interest can be concentrated prior to analysis.

Carbamates have not been previously screened due to the lack of methodology to analyse their formation. The carbamate modification is labile and therefore unable to withstand most conditions required for MS which remove the modification prior to identification (78). Recently a study has identified the formation of a carbamate on a whole peptide (Angiotensin) by using very gentle ESI technique (79) however this method would not be able to identify the site of carbamate formation due to the harsh conditions needed for ion fragmentation. The trapping methodology described in chapter 4 has removed this labile nature and allows for carbamate location analysis via MS.

It is proposed that understanding the regulation of phosphorylation in plants is key to understanding many plant functions (210). We hypothesise this is also the case for understanding the formation of carbamates.

The work described in this chapter uses the developed carbamate trapping method to produce robust covalently trapped carbamate sites on proteins. The trapping experiments were analysed using MS and the samples were run at the facility at

Durham University on an Applied Biosystems Qstar® XL mass spectrometer, however this machine has a limited sample rate compared to newer MS machines. A comprehensive review of proteins identified from an *E.coli* lysate compared two very similar MS machines; an Orbitrap and a Qstar elite. This study found that when analysing the same sample the Orbitrap outperformed the Qstar instrument, identifying over 2.5 times as many proteins and with higher mass accuracy than the Qstar machine (211). Samples were therefore also sent to the NUPPA facility at Newcastle University to be run on a Thermo LTQ (Linear Quadrupole) Orbitrap XL™ mass spectrometer.

The data collected from these machines was analysed with two different database software packages, GPM X!Tandem (Tandem) and MaxQuant, to identify protein carbamates.

5-3 Development of a method to identify unique proteins from a carbamate trapping reaction

5-3.1 Analysis of amount of protein extracted from *Arabidopsis* leaves

The samples were prepared for trapping by extraction of soluble protein from *Arabidopsis* leaves. Assaying the amount of protein in a plant sample is often difficult due to interference from phenols and chlorophyll and can lead to widely varying results (212). Bradford and BCA assays were carried out to calculate the amount of protein extracted from the leaves.

A Bradford assay works by measuring the binding of the Bradford dye which directly interacts with amino acid side chains, mostly arginine residues. This binding causes the transfer of a hydrogen from the dye converting it into the anionic form which causes a shift in the dye absorbance from 465 nm to 595 nm (213). This reliance on side chain groups can cause variety over different protein samples. A standard curve (Figure 5-2) was made using a range of BSA standard solutions and plotting the absorbance measurements. This graph can then be used to quantify the amount of protein in an unknown sample (213).

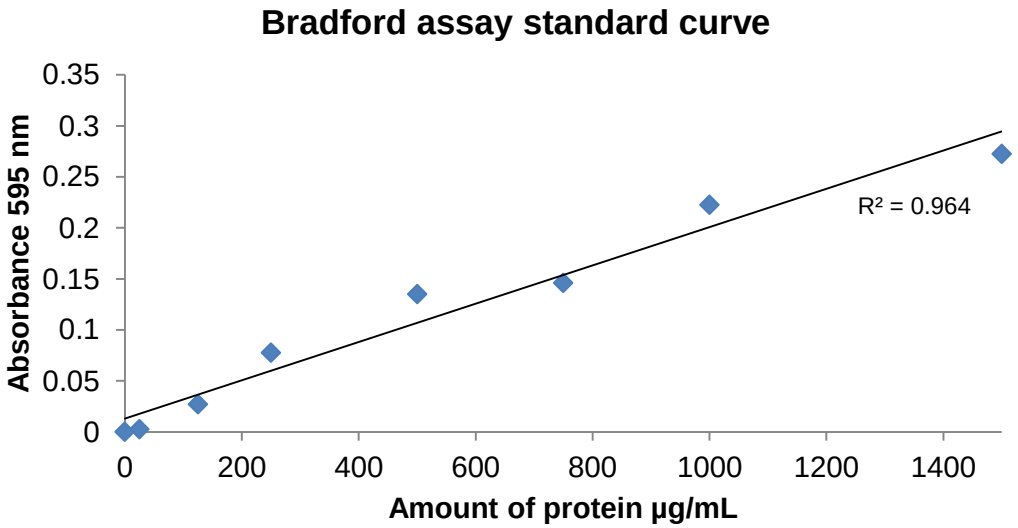


Figure 5-2 Standard curve from known concentrations of BSA using the Bradford assay, absorbance read at 595 nm. R squared values defines how closely the data fits the regression line.

$$y = 0.0002x + 0.0131$$

Equation 5-1 Standard curve equation from Figure 5-2.

Bradford assays are quick and easy to perform but can sometimes provide an underestimate due to interference by detergents and buffers binding with the dye (214) therefore a bicinchoninic acid (BCA) assay was also investigated.

A BCA assay works by the binding of copper ions (Cu^{2+}) to proteins in an alkaline environment. The peptide bonds within a protein reduce the Cu^{2+} ions from copper sulphate to Cu^{+}. The amount reduced is directly proportional to the amount of protein present in the solution. Then bicinchoninic acid chelates with these Cu^{+} ions forming a complex that is read at 562 nm. This method is known to be more tolerant to interference from detergents than the Bradford assay (215). A standard curve was produced in the same way using BSA standards (Figure 5-3).

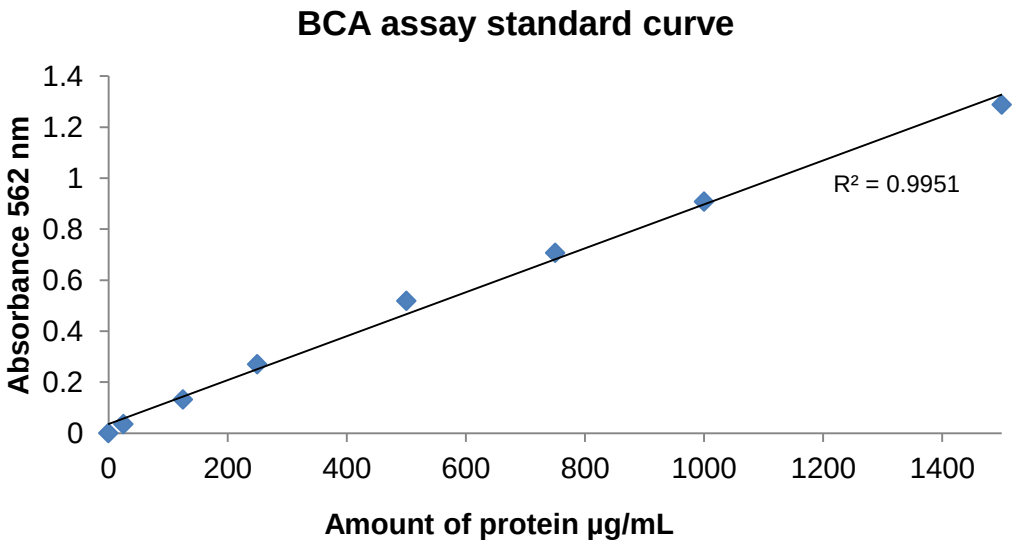


Figure 5-3 Standard curve from known concentrations of BSA using the BCA assay, absorbance read at 562 nm. R squared value close to 1 shows the data closely matches the regression line.

$$y = 0.0009x + 0.0358$$

Equation 5-2 Standard curve equation from Figure 5-3.

The two equations from these assays (Equation 5-1 and Equation 5-2) were then used to calculate the amount of protein in unknown samples.

5-3.2 Fractionation of the sample

Due to analysis limitations it is important to reduce the complexity of the sample before MS (202). Previous work investigating complex proteomes have used several methods to accomplish this.

As well as fractionation based on subcellular target location another method often used to separate proteins is by 2-DE (216). This method involves separating proteins on a gel by two dimensions, first by isoelectric point (pI) and then by molecular weight (201). This method was investigated in this work, however additional ethyl transfers from the TEO reagent to carboxylic acids containing amino acids (aspartate and glutamate) alters the pI of proteins and caused poor gel separation (data not shown).

A method often used in conjunction with 2-DE gels when investigating plant proteins is the removal of RuBisCO (217). As RuBisCO can account for almost 50 % of soluble leaf protein its removal can lead to an increase in identification of less abundant proteins (218). The removal of RuBisCO was achieved by using an anti-RuBisCO antibody packed into a spin column, this removes RuBisCO from the sample while allowing the rest of the proteins to flow through (217). The sample before and after passing through the RuBisCO removal column was run on an SDS-PAGE gel (Figure 5-4).

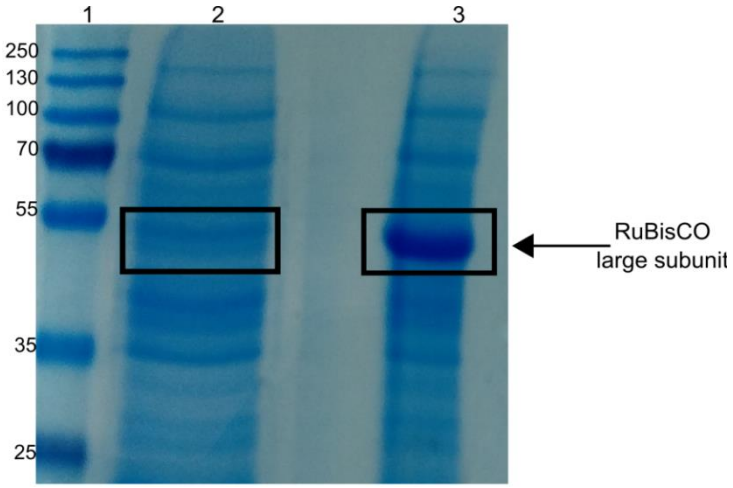


Figure 5-4 SDS-PAGE containing a MW marker (kDa) (lane 1) a RuBisCO depleted sample (lane 2) and the original leaf lysate (lane 3).

The SDS-PAGE compares a RuBisCO depleted sample (lane 2) to the original leaf lysate (lane 3). The RuBisCO removal can be seen by the absence of the large subunit in lane 2 at 53 kDa. The molecular weight is known by the use of a molecular weight ladder (lane 1) Figure 5-4.

Unfortunately, the capacity of the removal column (0.2 mg) was found to be too small to produce sufficient material for downstream analysis so this method was ultimately not suitable for reducing the sample complexity in this work.

Another method used to fractionate protein samples is by salt precipitation (219). This method works by increasing interactions between the proteins and the salt causing proteins to reveal hydrophobic residues and aggregate out of solution. The most commonly used salt is ammonium sulphate because of its high solubility. Fractions were taken at 20, 40, 60 and 80 % (v/v) salt and analysed by running on an SDS-PAGE gel (Figure 5-5).

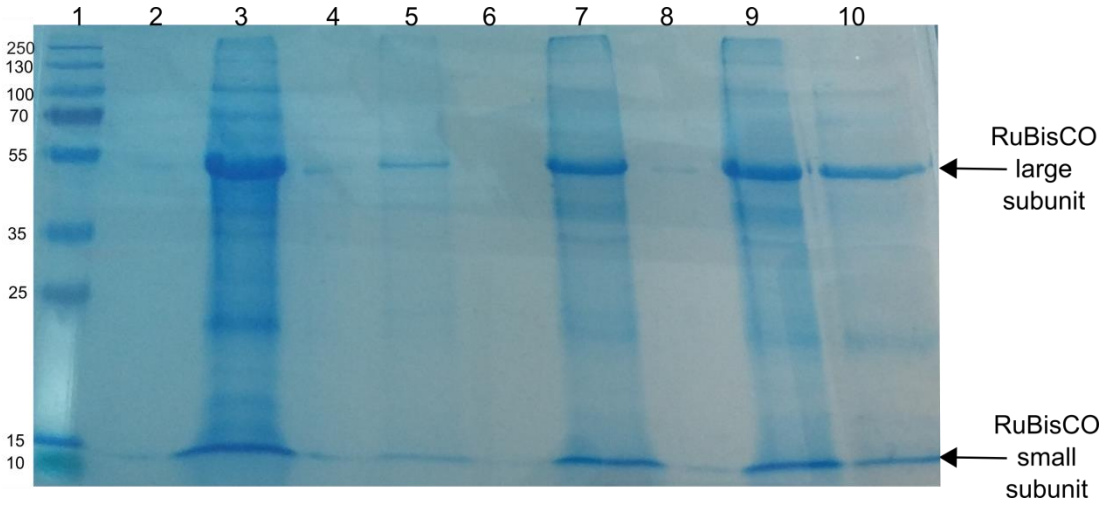


Figure 5-5 SDS-PAGE containing a MW marker (kDa) (lane 1), the original leaf lysate sample (lane 3) and four salt precipitation samples; F1 20 % (v/v) (lane 5) F2 40 % (v/v) (lane 7) F3 60 % (v/v) (lane 9) and F4 80 % (v/v) (lane 10). The RuBisCO large subunit (53 kDa) and small subunit (18.5 kDa) can still be seen in all lanes.

This method did not appear effective at fractionating the sample when visualised on a gel due to the main band in all the fractions still containing both RuBisCO subunits (large subunit, 53 kDa and small subunit, 18.5 kDa).

After the precipitation the samples were re-solubilised by dialysis into 50 mM phosphate buffer and were then used for individual trapping experiment and analysis by MS. Despite the appearance on the gel it was seen that there were unique proteins found in all fractions (Figure 5-6) with a total number of 91 unique proteins identified over the four fractions.

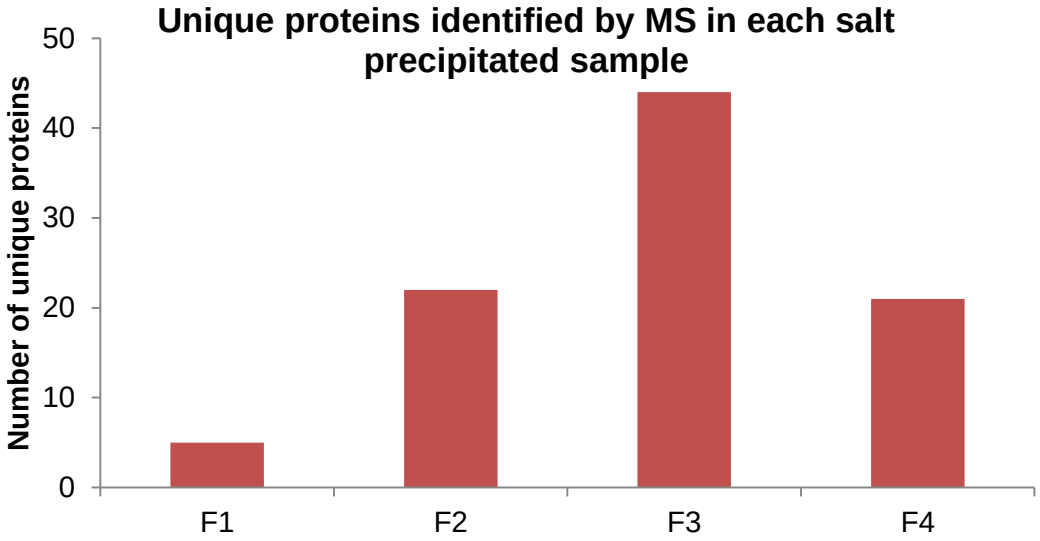


Figure 5-6 Number of unique proteins found in each salt precipitation sample, F1 20 % (v/v), F2 40 % (v/v), F3 60 % (v/v) and F4 80 % (v/v).

It could not be guaranteed that the proteins had not misfolded or lost their native state during the precipitation and re-solubilisation process; therefore, to ensure that the process of fractionating the sample did not interfere with the native state of the proteins, fractionation was carried out after the trapping experiment to provide confidence that the lysate was as close to physiological conditions as possible.

5-3.3 Solubilisation

During the carbamate trapping reaction there is additional transfer of ethyl groups by the TEO reagent to other amino acids. This can include any amino acid with free electrons but most commonly those containing carboxylic acid groups (aspartate and glutamate) (Figure 5-7) and containing amine groups (lysine and arginine). The transfer of ethyl groups blocks previous sites of hydrogen bonding by altering the charge state of the amino acid to neutral and causes protein precipitation during the trapping reaction.

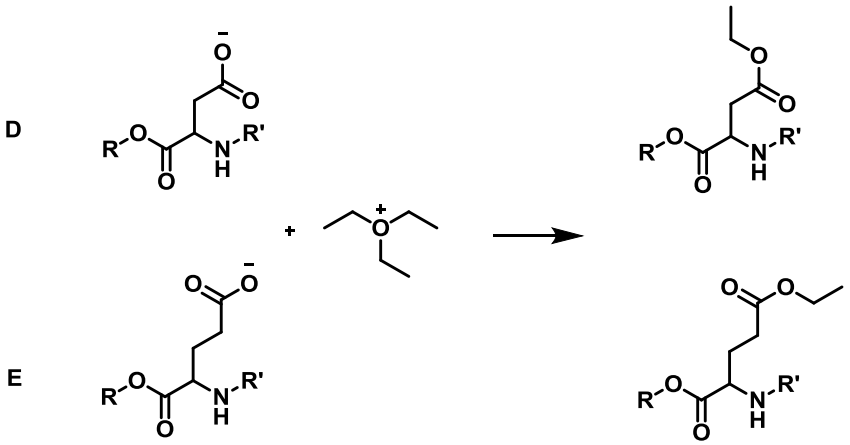


Figure 5-7 Ethylation of acidic amino acids (D and E) by TEO causing the alteration of negative charged side chains to neutral.

This precipitation causes denaturation of the protein, however this is not thought to be a problem for false positives because the protein structure is necessary to form a pKa environment capable of forming a carbamate.

The amount of protein present in the leaf lysate was compared to the amount of soluble protein present in the supernatant (SN) after trapping and a loss of ~70 % protein was seen (Figure 5-8). The precipitated proteins also cause difficulty for trypsin to digest the sample and therefore cause a reduction in the number of proteins identified by MS.

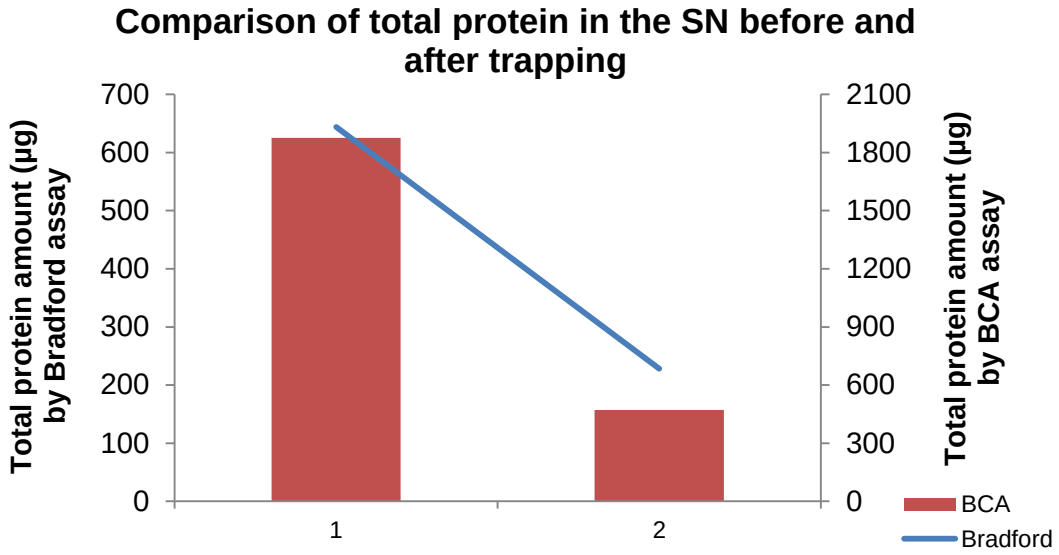


Figure 5-8 Total amount of protein present before the trapping experiment (1) and in the supernatant after the experiment (2) with the amount calculated using both Bradford (blue) and BCA assay (red).

The protein amounts recorded in Figure 5-8 were obtained using the Bradford and BCA assay equations (Equation 5-1 Bradford and Equation 5-2 BCA). The graph shows a very different amount of protein estimated depending on which assay was used (Bradford 644 µg to 228 µg and BCA 1876 µg to 470 µg). Previous work investigating both assays also revealed an underestimation by Bradford and an overestimation by the BCA assay (220,221).

Despite the exact amount of protein being inconsistent, both methods showed a drop (~70 %) in the amount of protein identified post-trapping with the lost protein contained in the precipitated pellet. To increase the amount of protein identified post-trapping, methods of solubilisation were investigated to increase the number of proteins in solution that could then be digested. These methods were based on previous membrane protein work as they are also hydrophobic proteins (222). The solubilisation methods investigated were; 0.1 % SDS, 10 % acetonitrile (MeCN) and heating at 95 °C and were compared to normal digest buffer as a control (50 mM ammonium bicarbonate (ABC) buffer) (223). The results of these solubilisation trials were analysed using both the Bradford and BCA assays on the supernatant after the

protein precipitate had been incubated for 1 h (2-3.9) and are shown in Figure 5-9.

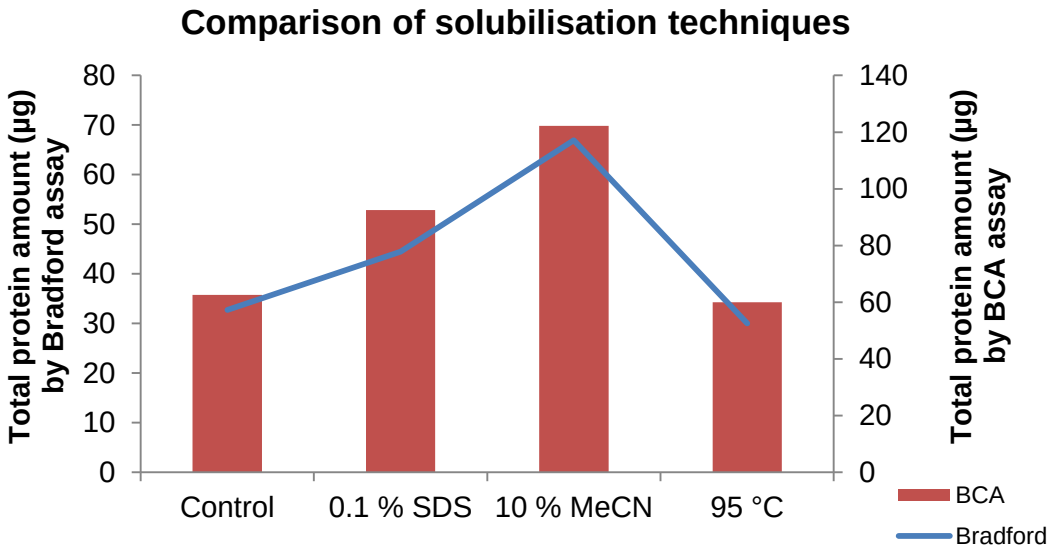


Figure 5-9 Total amount of protein from four solubilisation methods 4 solubilisation solutions 50 mM ammonium bicarbonate buffer, 0.1 % SDS, 10 % MeCN and heating at 95 °C) using both Bradford (blue) and BCA (red) assays.

The most successful condition was 10 % MeCN, a condition regularly used in digests as it is known to enhance the ability of trypsin (216). There are other detergents that are suitable for solubilising protein samples including Triton-X100 but they are not permitted for MS due to interference; they ionise very well and cause swamping of the protein in the sample (200).

Further solubilisation was sought by investigating the effect of increased buffer concentration. The concentration of buffer does not influence trypsin activity but can increase solubilisation of hydrophobic proteins by increased ionic strength. An increase of buffer concentration from 50 mM to 200 mM greatly increased the amount of protein seen by assay Figure 5-10 (the BCA assay was used).

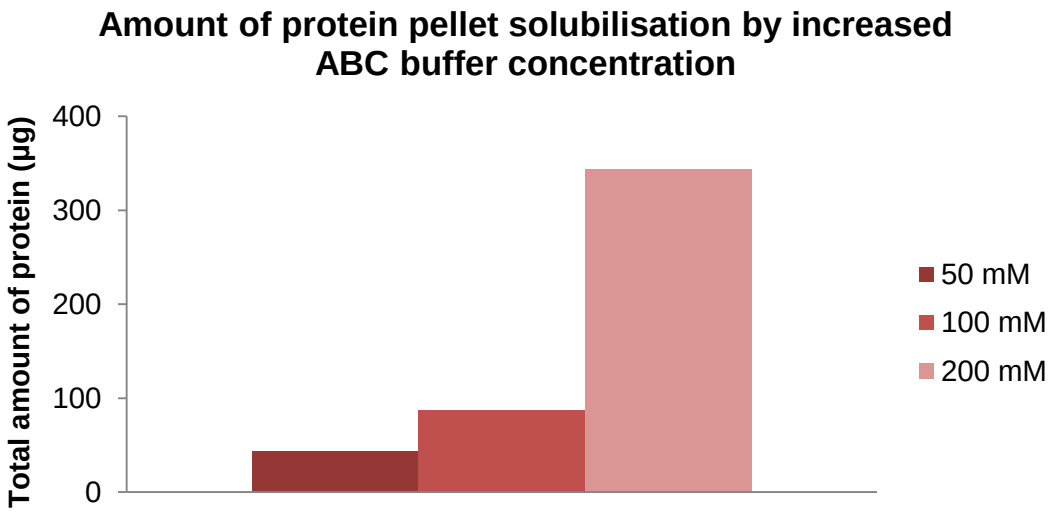


Figure 5-10 Comparison of precipitated pellet solubilisation with increased ABC buffer concentration by the BCA assay.

The increase of ABC buffer concentration from 50 mM to 200 mM increased the number of unique proteins identified from 44 to 344 therefore, digests were carried out using 200 mM ABC buffer containing 10 % MeCN (2-3.2). Fractionation was then carried out after digest to increase the number of identified peptides.

5-3.4 Fractionation of peptides

After sample digestion the method used for fractionation was via the use of a StageTip. A StageTip is a membrane inserted into a pipette tip over which the digested sample is then slowly passed (224), in this work a C_{18} membrane was combined with a cation exchange (SCX) membrane. The SCX membrane binds peptides based on their charge, the peptides are then fractionated by elution with increasing salt concentration (50 mM – 1 M) using ammonium acetate buffer (128).

The number of unique proteins identified using this fractionation method was far greater than compared to no fractionation (Figure 5-11).

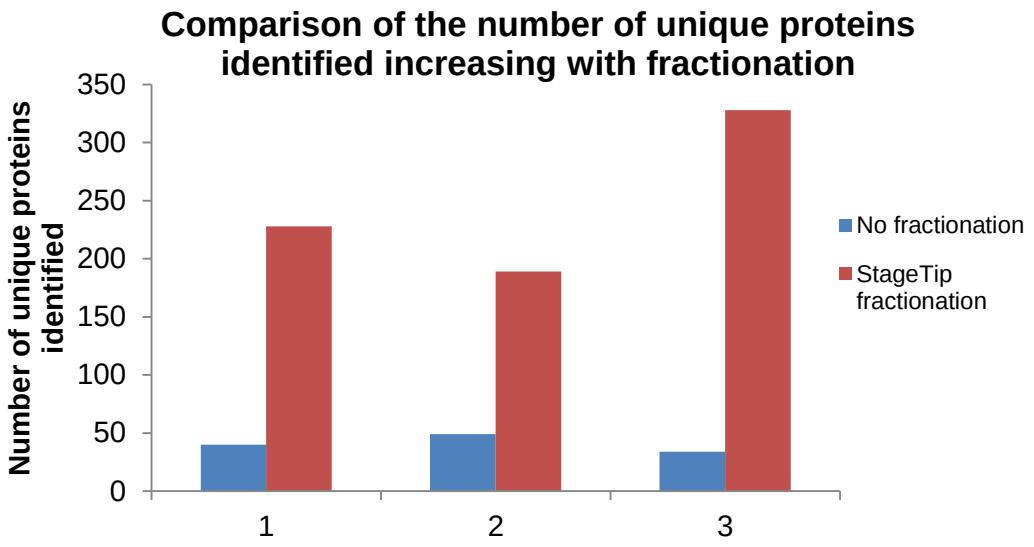


Figure 5-11 Comparing the number of unique proteins identified between three samples using sample fractionation.

The number of proteins identified was obtained using the analysis software Tandem (Tandem is explained further in 5-4.2).

Once the method development reached a stage capable of identifying ~300 proteins per sample a second MS machine was used to control for the introduction of variables by the machine. The NUPPA facility at Newcastle University contains an Thermo LTQ (Linear Quadrupole) Orbitrap XL™ mass spectrometer and so sample duplicates were compared between the two machines (Figure 5-12).

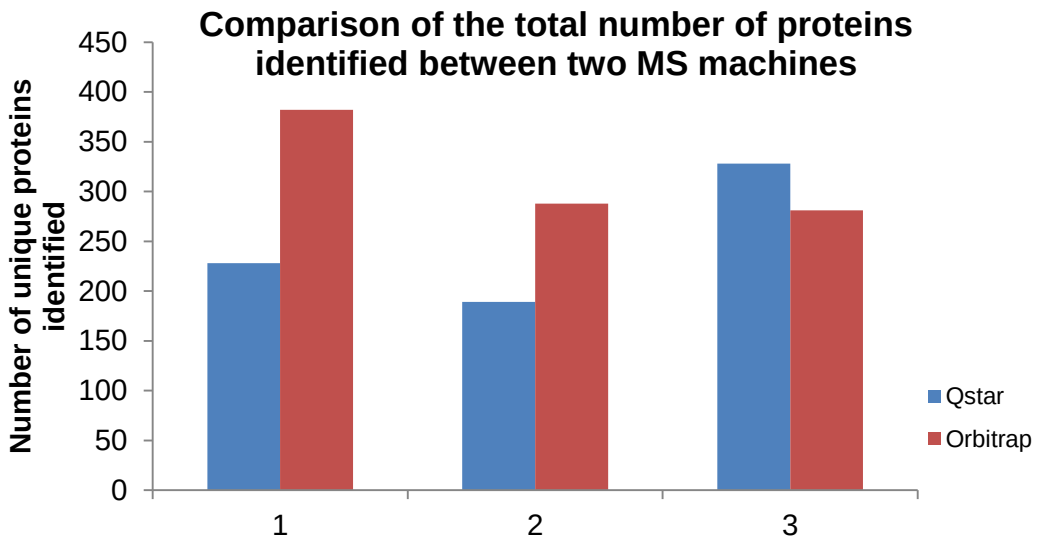


Figure 5-12 Comparison of the number of unique proteins identified between the Qstar and Orbitrap MS machines.

Three different experimental samples were run at both facilities. The Qstar samples were fractionated by StageTip into 5 fractions before ESI-MS. The Orbitrap samples were fractionated on an in-line SCX column before injection. A two sample t-test showed no significant difference between the results from these samples.

5-4 Analysing protein carbamates

5-4.1 Difficulty in analysing data

MS of a complex protein mixture produces a large amount of MS-MS data. This data is analysed using software which assigns proteins from MW and fragment mass information (225). This is performed by comparing *in silico* protein sequences with experimental data, these matches are then scored by the software. PTMs greatly increase the complexity of mass-based searching by increasing the search variables (226). The search time for comparison of spectra increases dramatically with every newly possible modification site (227). Identifying PTMs is very important for understanding protein function but many modifications are still not present in standard databases. Previous studies investigating PTMs have used software such as FindMod which allows the searching of PTMs, but contains a limited list of modifications to be searched (228).

There are many available online software options to analyse ESI MS data, however most are limited to the search of two variable modifications and do not permit user-defined modifications (the addition of unknown modifications). This limitation in software is especially important in this work as most software modification tables only permit ethylation to occur on amino acids aspartate and glutamate, but in this work ethylation can occur on a variety of other amino acids. As well as the ethyl group modification the modified carbamate group also needs to be searched for. Therefore, only software that permits multiple variable modifications, and that can be user-defined, can be used for this analysis. The amino acids suggested to be ethylated in this work as well as the two sites of carbamate formation (lysine and N-terminal) are detailed in Table 5-1.

Table 5-1 Variable modifications that can occur during the reaction with TEO.

Modification	Amino acid	Modification	Amino acid
Ethylation	D	Carbamate	K
Ethylation	E	Carbamate	N-terminal
Ethylation	K		
Ethylation	R		
Ethylation	H		
Ethylation	T		
Ethylation	S		

In order to reduce the chances of finding a false positive the analysis searches were carried out only including ethylations on D and E amino acids. This may have meant that a possible carbamate would have been missed (false negative) but gives greater confidence to the hits found. The other listed amino acids in Table 5-1 are amino acids with free electrons that could possibly be ethylated due to the strong ethylating nature of TEO.

This level of modification is unusual for a proteomics experiment and therefore requires specifically tailored search software. Two software packages were used for the analysis of the data produced in this work.

5-4.2 Software used for analysis

All MS analysis software have similar initial settings for searching data, these include the protease used for the digest, the type of MS machine the data was collected on, the species database to search and any chemical modifications carried out on the sample (229). However, software packages then differ in the way they score matching data and the PTMs they are capable of searching for.

Mascot is a very popular MS search software. Mascot is currently the most common software for peptide identification (197) and works based on a probability scoring system. This system works by calculating the probability that the experimental data could match a theoretical spectrum randomly, if the peptide is correct the probability of this being a random match will be very small due to the number of matched

peaks. The online version only allows for searching of four variable modifications due to the increased chance of a random hit with each additional modification (226). An additional disadvantage of Mascot is that the freely available version cannot accept user-defined PTMs (197). Mascot was therefore not used for the analysis of this data.

GPM X!Tandem (Tandem) is an online software package specifically adapted to handle large amounts of data. This is achieved by a two-step process, the software first identifies a set of proteins based on abundant peptides and narrows the specified search database down from a whole genome to this subset, this new subset is then more carefully refined (197). This software is capable of accepting user-defined PTMs and due to its fast analysis rate was the first software used here for MS-MS analysis. The Tandem scoring system is a matched peak score instead of probability based and is carried out based on knowledge of amino acid intensity patterns upon fragmentation. These theoretical spectra are then compared to the raw data and the shared peaks between the spectra are used to calculate a score (230). Tandem reports e-values to show how unlikely it is for the assigned peptide to be matched to the spectrum by chance. A reverse database is also searched and compared to look at random chance matches under the thresholds used for searching.

A study recently compared Tandem to Mascot based on the number of false positive assignments made and showed both to have 99 % specificity but Tandem outcompeted Mascot with a higher sensitivity of 74 % to 71 % respectively (231). As with all analytical searches it was important to guard against false positives to increase the confidence in the carbamate hits found. Use of a target-decoy search strategy by using a reverse database to create random search sequences was a procedure shown to give reproducible results for the same input data using both software (231) and is used in all searches performed using this data.

An important internal marker for hit confidence was present within the digest reaction. The protease used for digesting the sample, trypsin, cleaves at the carboxyl side of lysine and arginine residues. The trypsin catalytic pocket contains an aspartate residue which is stabilised by the positive charge on lysine and arginine residues during the cleavage reaction, and mutation of this residue to a

positive amino acid removed all catalytic activity (232). Therefore, as the addition of a trapped carbamate on a lysine converts its charge state to neutral and adds additional bulk, trypsin cannot cleave at this site. Missed cleavages are rare for trypsin so a carbamate suggested on a missed cleavage adds confidence to the presence of that carbamate. This process of searching for missed cleavages has previously also been implemented while looking at acetylation on lysine groups (233). For this reason, only carbamates found on miscleaved lysines were considered to be potential hits.

The Tandem software has two disadvantages. The first is that it does not display the raw data; the software displays a user interface image of the fragmentation showing the matching ions colour coded. This is the software interpretation of the data but doesn't allow the user to check the raw data for background level or ion intensity or the fragmentation pattern surrounding the modification. The other disadvantage is related to its ability to search data so quickly. The first step of the search involves narrowing the search database by creating a smaller one from the data. This requires the presence of an unmodified peptide for the protein to be included in the smaller made subset search database. In the case of these experiments the high level of modification could mean that some proteins were not considered due to not having an unmodified peptide present in the sample. Therefore, further analysis of all samples was carried out using another software package, MaxQuant, to compensate for any potential deficiencies in Tandem.

MaxQuant is standalone software which also allows for user-defined modifications; however, unlike Tandem it searches all of the data. This greatly increases the analysis time but also reduces the risk of missing a modification because the protein was not selected during the first analysis step. MaxQuant combined with the search engine Andromeda uses a probability search algorithm like Mascot (130) that matches peptide fragment information to generate spectra based on probability.

MaxQuant is proposed to obtain a higher level of accuracy based on its method of assigning peak m/z values, whereby the software fits each peak with a Gaussian distribution and the centre of this distribution becomes the mass estimate for the peak. In this way it determines masses with a far higher accuracy than Mascot (234). The visualisation of data from MaxQuant also allows the user to see the raw

data spectrum and can be used to combine multiple data samples to increase scoring.

5-4.3 Carbamate hits found

A carbamate hit was searched for in Tandem or Maxquant by searching for the mass of a carbamate with the addition of an ethyl group (72.0211 Da, Figure 5-13).

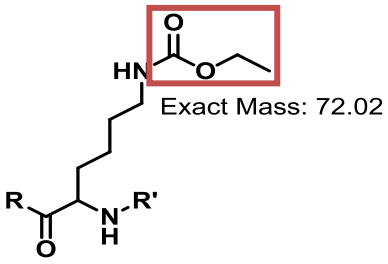


Figure 5-13 Exact mass of carbamate addition on a lysine side-chain.

5-4.3.1 Tandem

The Tandem software identified several carbamates on lysine residues at previously unknown sites. Some of the most promising hits are described below and a summary table of other hits is displayed in 5-4.5.

The graphical representation of the results from Tandem shows the sequence of the identified peptide with the modified residue highlighted in blue and the modification label shown underneath (carboxyethyl). The calculated peaks that would arise from the fragmentation of the peptide are shown (on the right) with those that match the spectrum highlighted in red which are the matches used to calculate the log e value score (Figure 5-14).

In order to be confident that the newly found carbamates are true PTMs there were several criteria that were put in place. Both analysis software packages provide confidence scores for peptide assignment. The Tandem software works by calculating a shared peak count (SPC) between the theoretical spectra and the raw

data spectra. A confidence score is then calculated based on how close the spectrum match together.

At2g45180 Lipid transfer protein (LTP)

A carbamate hit was identified at K44 on the LTP found using both the Tandem and MaxQuant software. After the first hits were identified experiments were repeated using ^{13}C labelled sodium bicarbonate to demonstrate that the carbon dioxide binding to the protein is caused by the reaction conditions. This modification could be searched for by altering the user defined modification from 72.0211 to 73.0244. A matching ^{12}C and ^{13}C hit was found for LTP K44 which greatly improved the confidence in the modification (Figure 5-14).

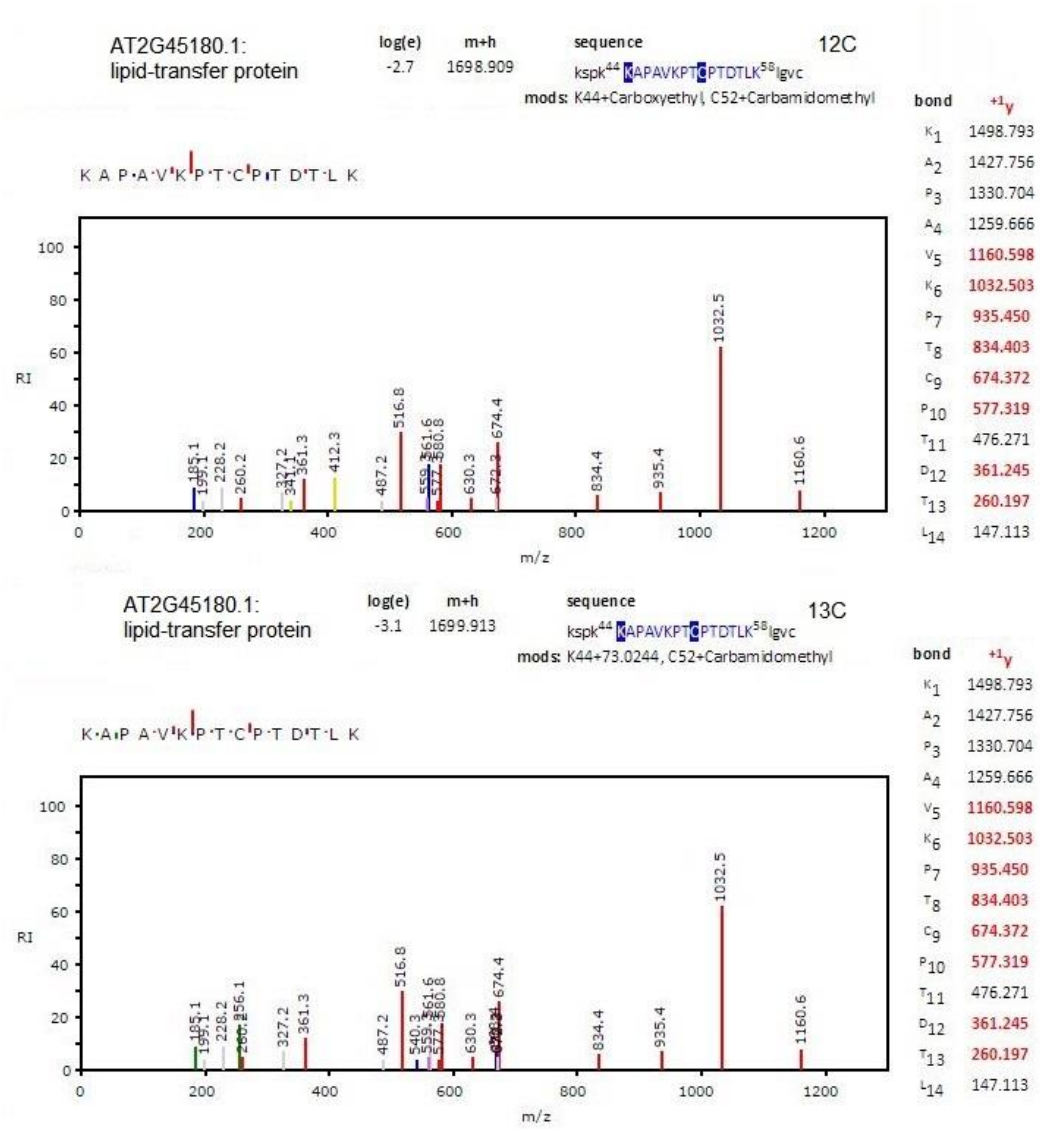


Figure 5-14 ^{12}C (top) vs ^{13}C (bottom) carbamate trapped on K44 of the lipid transfer protein.

This successful experiment shows that the carbamate is forming under the reaction conditions. The LTP is present in the chloroplast thylakoid membrane which is the centre for oxygenic photosynthesis (235) and has been implicated in the stress response (236).

Atcg00490 RuBisCO

A carbamate hit was identified at K183 in repeated samples on RuBisCO. These hits were found with high confidence scores (−5.1 from Tandem, 170 from MaxQuant). This hit at K183 does not match the already known RuBisCO carbamylation site at K201 (Figure 5-15).

The Tandem software scores a 'hit' using an e value which is a score comparing the chances of the hit being a false positive. For example, a score of −5.1 means there is a 0.0000051 chance that the hit is a false positive.

RuBisCO is already known to interact with CO_2 as the site of carbon fixation in the Calvin cycle. The site of carbamate formation already known in RuBisCO is K201 (96). This site is within the same RuBisCO protein region which is conserved amongst higher plants as the trapped K183. The protein sequences were compared between the Spinach of the previous Lorimer, 1983, K201 conclusion and the Arabidopsis used within this work and the sites are definitely distinct.

This carbamate site is a non-exchangeable site due to the stabilisation added by the Mg^{2+} ion. The trapping methodology developed here is able to identify exchangeable sites that have previously not been identified. Therefore, we hypothesise that K183 is an exchangeable CO_2 binding site on RuBisCO.

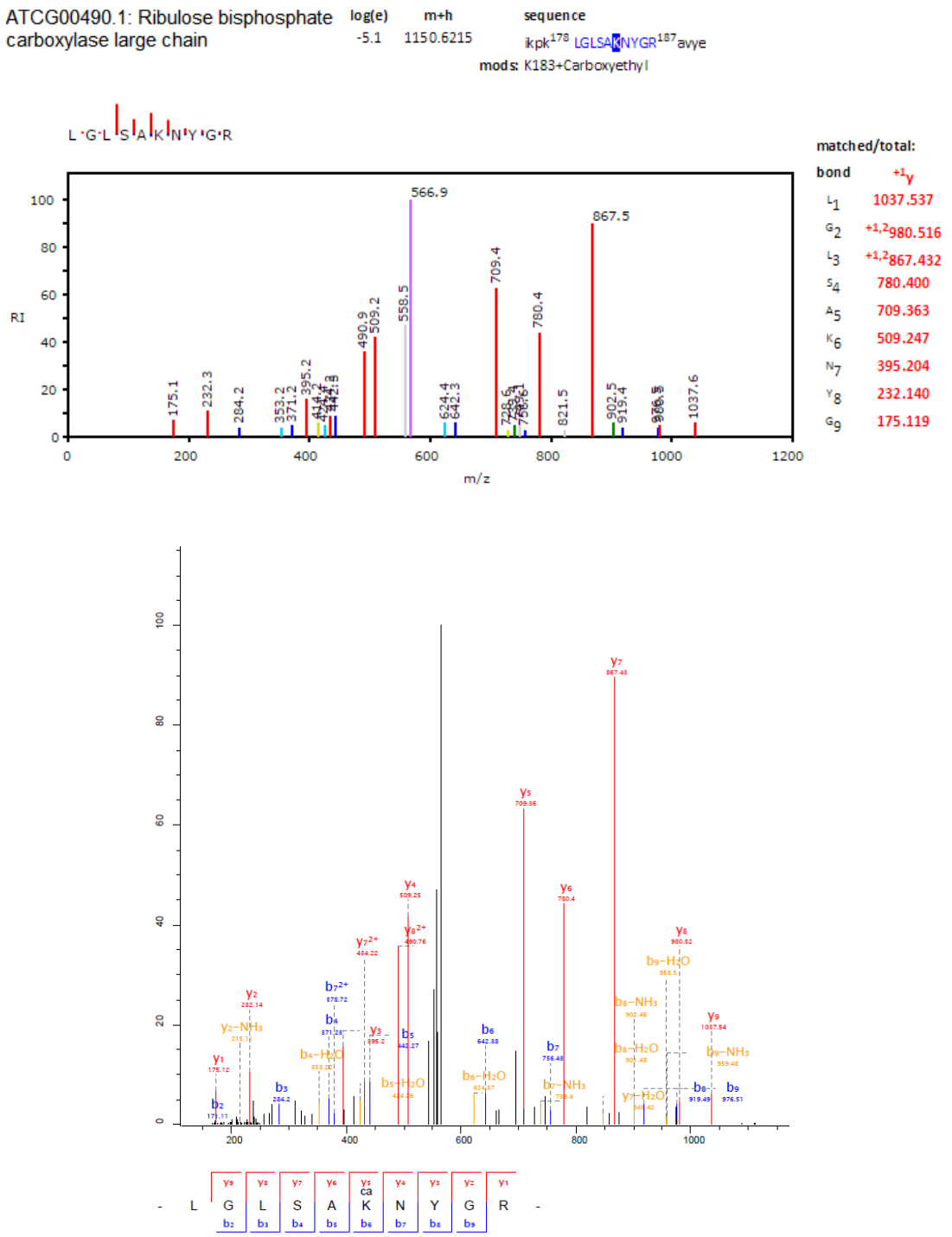


Figure 5-15 Software displays of the K183 carbamate hit found on RuBisCO using both Tandem (top) and MaxQuant (bottom).

At3g32980 Peroxidase

A carbamate was identified at K262 on Peroxidase. This site was seen using both software and had the best score found in the Tandem searches (−7.6) (Figure 5-16). Examination of the spectrum reveals this is because there are many peaks that match the peptide fragmentation patterns.

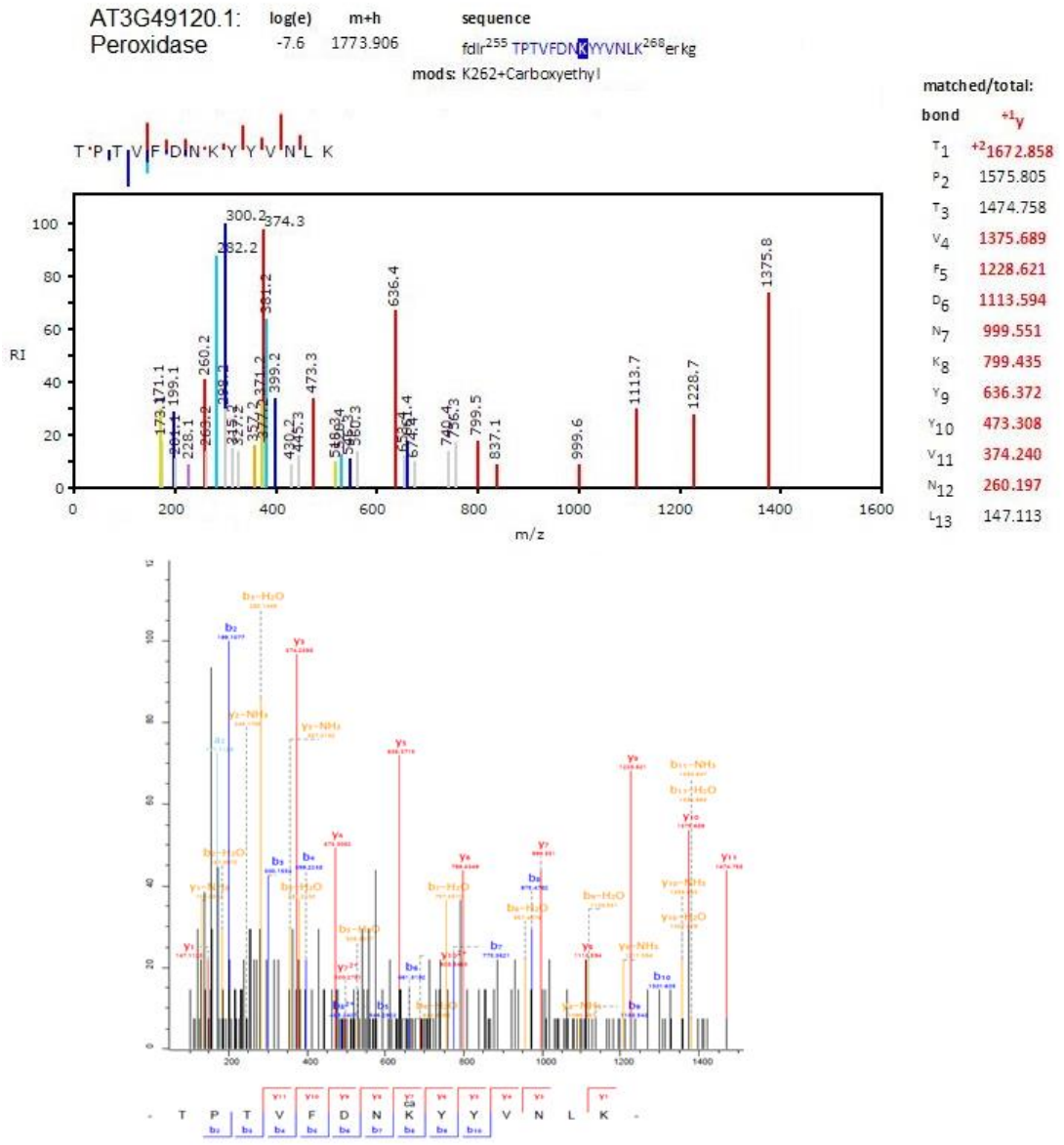


Figure 5-16 Software displays of the K262 carbamate hit found on Peroxidase using both Tandem (−7.6, top) and MaxQuant (bottom).

It has previously been observed that increased levels of O_3 cause enhanced activity of peroxidases but that this difference was seen to change under high levels of CO_2, implying a possible effect on peroxidase by the CO_2 present (237).

At2g21330 Fructose bisphosphate aldolase 1 (FBA1)

A carbamate was identified at K293 of FBA1 in repeat samples and with a confidence score of −5.8 (Figure 5-17).

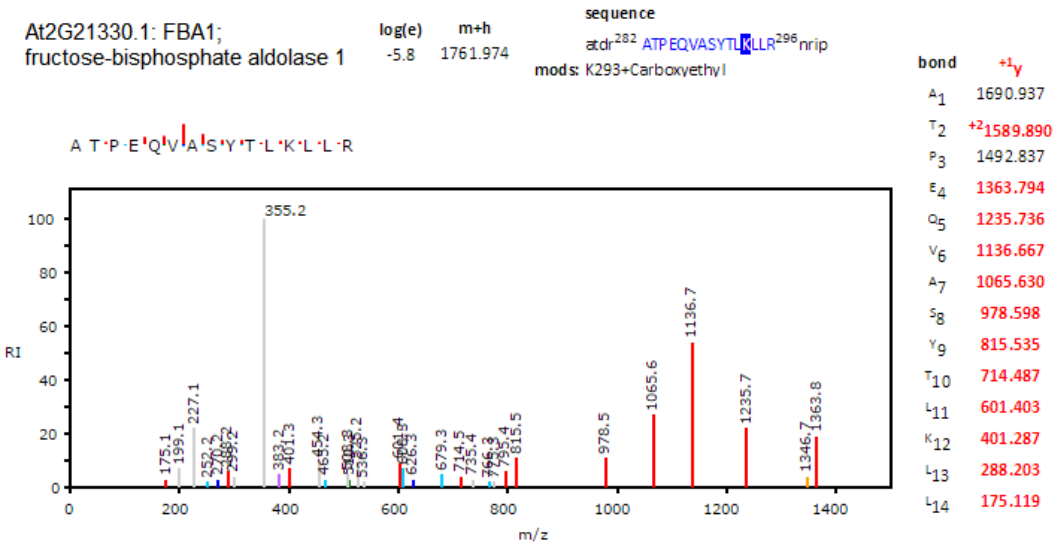


Figure 5-17 Tandem result showing a trapped carbamate at K293 on protein FBA1 with a confidence value of −5.8.

Fructose bisphosphate aldolase has been identified as a redox sensitive target responding to ABA signalling in guard cells (238).

5-4.3.2 MaxQuant

The scoring system for MaxQuant differs from Tandem by being based on a probability scoring strategy similar to Mascot.

The results displayed by this software allow a clearer look at the fragmentation pattern seen from the raw data by displaying the peptide sequence assigned and showing the spectrum peaks and where matches occurred. The ability to visualise the raw fragmentation in this way rules out false positives that did not have fragmentation around the carbamate modification.

The MaxQuant software also identified several carbamate hits, results were found that matched the peroxidase K262 and the RuBisCO K183 but also new protein hits.

At4g21280 Photosystem II subunit Q

A carbamate was identified at residue K109. This site was found with a high score and a good fragmentation pattern (Figure 5-18).

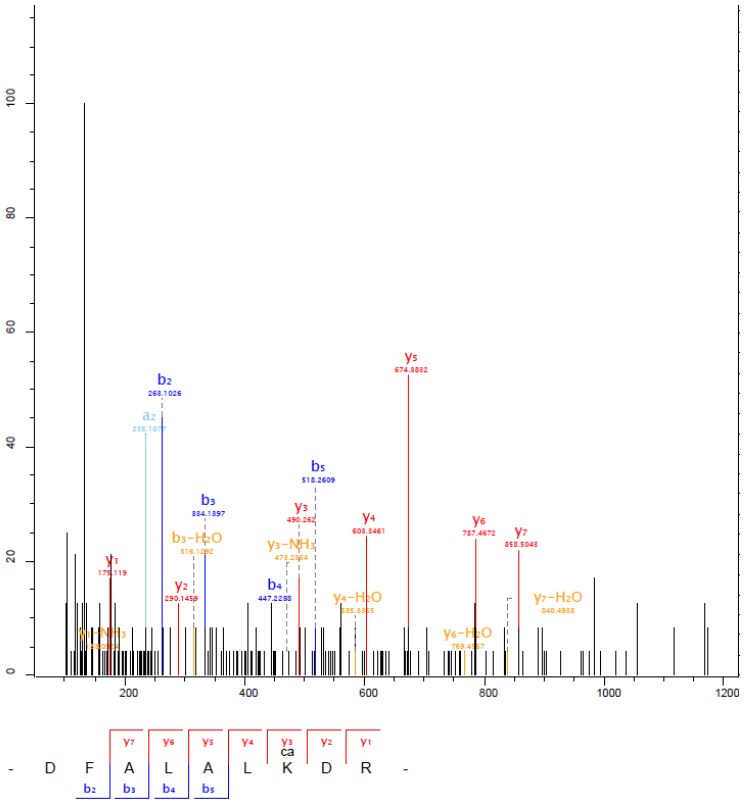


Figure 5-18 Fragmentation pattern of the peptide containing the K109 carbamate on PSII subunit Q.

Figure 5-18 shows the fragmentation occurring around the carbamate modification which improves the confidence in the presence of the carbamate.

Subunit Q is a subunit of PSII involved in oxygen evolution (239) and it has been previously established that PSII activity requires high amounts of dissolved CO_2 (240).

At1g08070 Tetratricopeptide repeat containing protein

A carbamate was identified at residue K696 of a tetratricopeptide repeat containing protein. This hit has a high score of 137.9 (Figure 5-19) and was identified in multiple experiments.

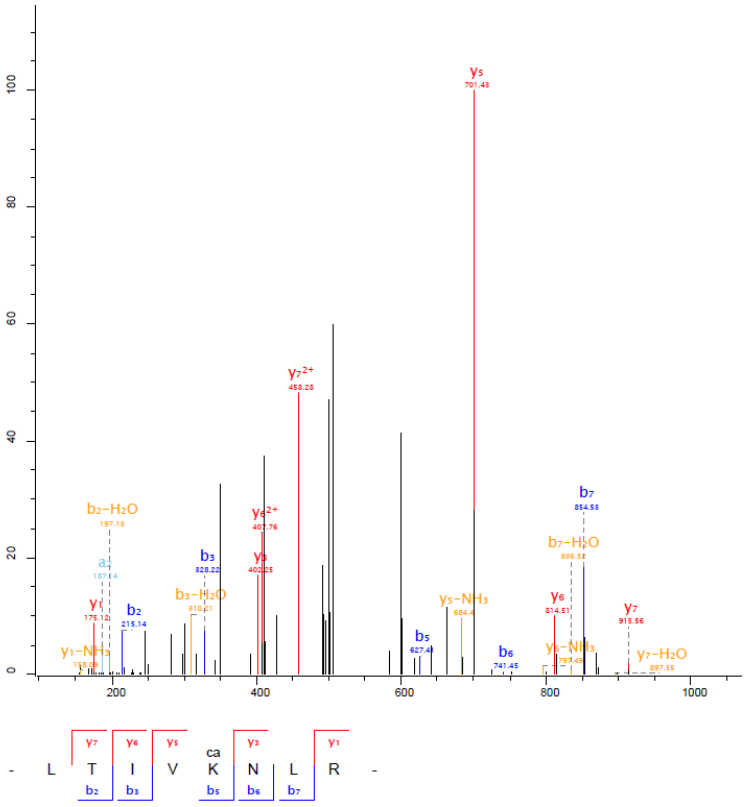


Figure 5-19 Fragmentation pattern of the peptide containing the K696 carbamate on Tetratricopeptide repeat containing protein.

This protein encodes a chloroplast RNA editing factor (241) and is involved in RNA editing of genes from plastids which encode many genes for photosynthesis.

At4g02790 GTP-binding family protein

A carbamate hit was identified at residue K207 on a GTP-binding family protein. Toc complexes are located in the chloroplast outer membrane and together with TIC complexes (inner membrane) facilitate the transfer of proteins out of the chloroplast

(242). Proteins in the Toc complex are responsible for chloroplast biogenesis and are GTP-binding dependent so require the activation of GTP-binding proteins (243).

5-4.4 False positive examples

As described above, the best way to improve confidence in a hit is for it to be present as a missed trypsin cleavage site and for the ESI fragmentation pattern to surround the site of modification. These are the best ways to rule out a false positive. Examples of potential false positives found by the software are shown below.

At5g55560 Protein kinase superfamily protein

A carbamate hit was purportedly found on a protein kinase superfamily protein at residue K187. However the site of modification was at a cleaved lysine residue so it is unlikely that a carbamate could have been attached here (Figure 5-20).

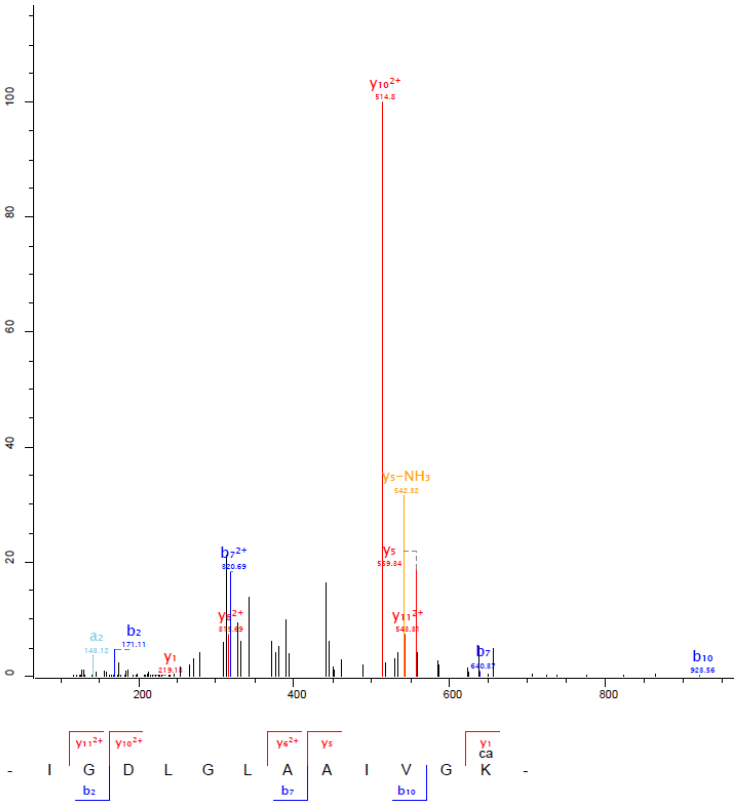


Figure 5-20 Example of a false positive carbamate site at a cleaved lysine residue.

Observing the raw data also shows a very poor fragmentation pattern for the whole peptide.

At1g12800 Nucleic acid-binding protein

Some sites of suggested carbamates had high scores but when the raw data fragmentation pattern was observed the fragmentation was not present around the modification. An example included here was at residue K82 on Nucleic acid-binding protein and came with a high score of 104 but the fragmentation pattern could not be used to ascertain whether the carbamate was genuine (Figure 5-21).

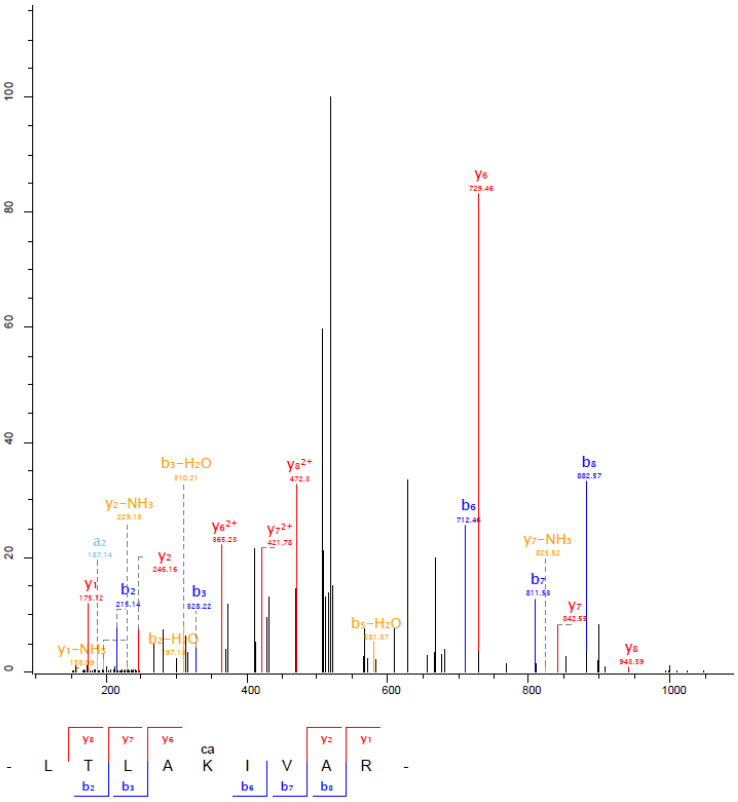


Figure 5-21 An example of a bad fragmentation pattern not around the carbamate modification.

5-4.5 Summary of total hits

A summary of the other hits found by both software can be seen in Table 5-2 including the protein, residue containing the carbamate and the software score.

Table 5-2 Combined carbamate hits found using both analysis software from data run on both the Qstar and Orbitrap MS machines.

MS run on	Software analysed with	Genome number| Protein identification	Residue	Score
Qstar	Tandem	Atcg00490| Rubisco	K183	-2.2
Qstar	Tandem	At2g42530| Cold regulated 15b	K107	-2.0
Orbitrap	Tandem	At1g20630| Catalase	K574	-1.2
Orbitrap	Tandem	At4g38970| Fructose bisphosphate aldolase 2	K292	-2.1
Orbitrap	Tandem	At5g60390| GTP binding elongation factor	K164	-1.6
Orbitrap	Tandem	At3g12780| Phosphoglycerate kinase	K378	-1.1
Orbitrap	MaxQuant	At3g52590| Ubiquitin extension protein	K48	92.19
Orbitrap	MaxQuant	At2g44500| O-fucosyltransferase family protein	K307	89.89
Orbitrap	MaxQuant	At3g54400| Eukaryotic aspartyl protease	K251	73.23

Orbitrap	MaxQuant	At1g17860| Kunitz family trypsin inhibitor protein	K178	57.17
Orbitrap	MaxQuant	At3g16850| Pectinlyase-like superfamily protein	K132	72.79
Orbitrap	MaxQuant	At5g36700| 2-phosphoglycolate Phosphatase1	K123	50.31
Orbitrap	MaxQuant	At5g23600| RNA phosphotransferase	K70	85.21
Orbitrap	MaxQuant	At2g35140| Development and cell death domain	K168	43.77
Orbitrap	MaxQuant	At3g26240| Cysteine/histidine rich domain family	K317	83.87
Orbitrap	MaxQuant	At1g31730| Adaptin family protein	K63	88.50

5-5 Conclusion

This chapter investigated several methods of sample fractionation to increase the number of proteins identified and therefore the chances of finding a carbamate. Methods were also developed to improve solubilisation of the sample after the trapping experiment. These advances combined to improve the number of unique proteins identified from 91 to 382 over the course of this work.

Fractionation methods were investigated to increase the number of unique proteins identified within the complex sample, with a maximum number of 382 identified in the final protein screens. This is still significantly lower than ~4000 proteins described in several literature studies.

Different analysis software were investigated to allow the use of multiple user–defined modifications and to visualise the raw data. Using these two software several novel carbamate sites were deduced with the use of a decoy database and by searching for only miscleavage events. These two conditions combined with the software scoring system and repeated hits in multiple samples gave several very confident carbamate sites (5-4.3.1, 5-4.3.2).

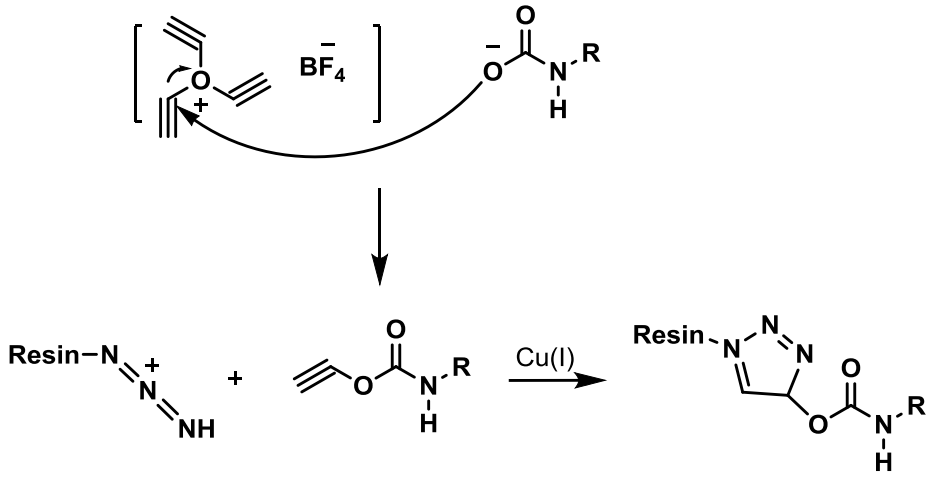
This chapter has described the development of a method to trap previously unknown carbamates within a complex protein system. This method has been found to be capable of trapping both soluble and insoluble proteins shown by the trapping of a carbamate on the membrane bound protein LTP.

5-6 Future work

The biggest limitation in this work is the coverage of the proteome, further work needs to be performed to increase the number of proteins identified. In previous investigations that managed to identify thousands of proteins the main difference is much more extensive chromatography for fractionation of the sample.

Another method to increase the chances of finding a carbamate modification would be to create an enrichment method. Another common lysine PTM is acetylation, this modification is often enriched for by use of an anti-acetyl-lysine antibody (244). A carbamate specific antibody could be raised against synthetically made carbamate trapped acetyl-lysine (chapter 4) bound to BSA as the antigen.

Other enrichment strategies utilised for phosphorylation screening involve a pull down or chromatography method. This style of enrichment would be possible in this work by synthetically altering the TEO trapping reagent to be able to click attach to a purification column. The method suggested here would be to alter the ethyl groups to alkyne groups which could be clicked to azide groups attached to a resin (Scheme 5-1).



Scheme 5-1 the process of using alkyne-azide click chemistry to modify the TEO trapping reagent so be used to enrich the sample on an azide resin.

The next step for work in this investigation is the validation of a newly discovered carbamate hit on a previously unknown site. The protein chosen for investigation was FBA1 due to the high confidence score given and that it was found in multiple samples using both analysis software. The process of protein production and validation of activity by assay is discussed in chapter 6.

Chapter 6: Expression and validation of FBA1

6-1 Overview

Proteomic analysis of an *Arabidopsis* leaf lysate revealed several new sites of carbamate formation (chapter 5). Efforts were made during the search to reduce the chances of a false positive occurring, however confirmation that the carbamate is forming on an isolated and purified protein would provide validation for the mass spectrometry data.

The protein chosen for validation was fructose bisphosphate aldolase 1 (FBA1). FBA1 was chosen as it was identified in repeated experiments with a high confidence score. Also, for practical reasons, FBA1 is not a membrane protein and very similar proteins have been expressed as recombinants (253).

FBA1 is a fructose bisphosphate aldolase located in the chloroplast. FBA1 is one of the enzymes of the Calvin cycle where it is involved in the condensation of F-1,6-BP (254) as well as playing a key role in glycolysis and gluconeogenesis (255). FBA has been shown to respond to ABA and ROS which indicates its importance in abiotic stress responses including responses to CO_2.

The work described in this chapter is the process of producing recombinant FBA1 isolated from BL21 *E. coli* cells and purified using affinity column chromatography. This purified protein was then assayed to demonstrate its activity before being used for a trapping experiment to confirm carbamate formation. To investigate the site of carbamate formation a mutant FBA1 K293A was also produced converting the carbamate site lysine to an alanine to investigate any activity changes.

6-2 Factors in recombinant protein expression

6-2.1 Choice of organism

Escherichia coli (E. coli) is a suitable host organism for recombinant protein expression due to its fast doubling time and its ease of transformation with exogenous DNA (256). Protein expression using *E. coli* is a well-established procedure (257).

There are many different *E. coli* cell strains. This work utilised the DH5α strain for DNA production and the BL21 strain for protein production. DH5α cells contain a *recA* mutation that reduces recombination of DNA and an *endAI* mutation which reduces plasmid digestion (258). The *E. coli* strain BL21 is deficient in the Lon and OmpT proteases which degrade many foreign and extracellular proteins (256) making them suitable for producing recombinant proteins.

6-2.2 Choice of cloning vector

A cloning vector is the plasmid used to insert the desired DNA into the bacterial cell. A cloning vector has three important criteria: 1) an origin of replication (*ori*), 2) a selection marker and 3) restriction enzyme digest sites.

In this work pET-14b was used as an expression vector (Figure 6-1). This vector system contains the pBR322 *ori* which is the site at which replication of the plasmid is initiated.

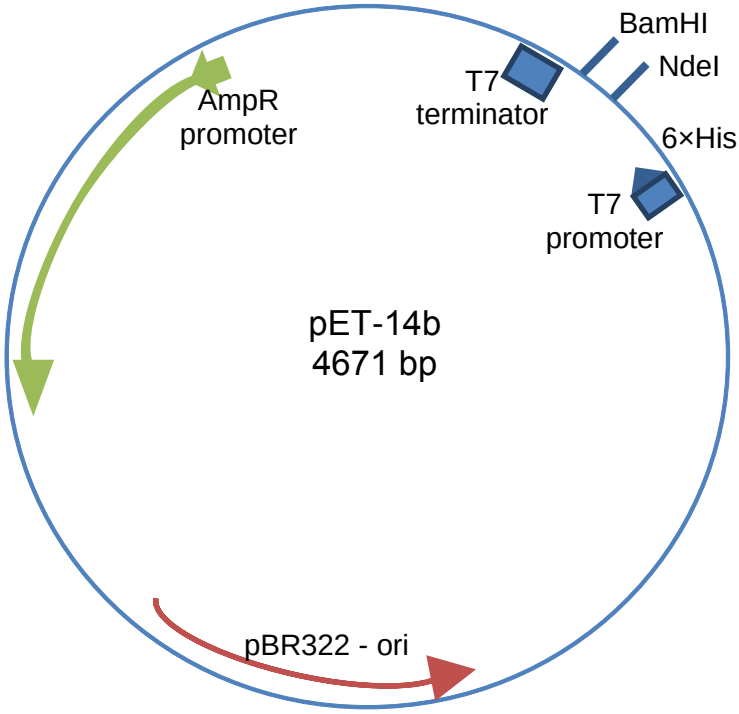


Figure 6-1 pET-14b expression vector containing AmpR antibiotic resistance marker, pBR322 ori and T7 promoter sites.

In order to control protein expression the pET vector system contains an antibiotic resistance gene, the *lacI* gene from the lac operon that codes for the lac repressor (LacI) and the gene of interest (GOI) inserted after the T7 promoter DNA sequence. This T7 promoter prevents leaky expression by only matching T7 RNA polymerase (RNAP) from T7 bacteriophage. LacI binds to the lac operator and prevents transcription. The LacI repressor is removed from the operator of the lac operon in *E. coli* by allolactose allowing transcription. Artificial induction of the system in the pET vector is achieved through the use of a non-hydrolysable analogue, IPTG, which is capable of binding the lac repressor but is not metabolised within the cell (259).

To ensure that the bacterial cells have taken up the plasmid containing the GOI, a selection marker is expressed from the plasmid in the form of an antibiotic resistance gene. The antibiotic resistance used in this DNA manipulation was ampicillin resistance introduced by the *amp*^{r} gene. *Amp*^{r} encodes for the β-lactamase enzyme which hydrolyses the β-lactam ring of ampicillin and removes its ability to prevent bacterial cell wall biosynthesis (260).

6-2.3 Purification tags

In order to purify the protein of interest (POI) from the growth culture, a tag is attached so that purification can be carried out in one-step by affinity chromatography. The tag is encoded for by the vector used and in pET-14b is a His-tag which allows the protein to be purified using a Ni^{2+} charged resin with the protein being eluted using increasing concentrations of imidazole for competitive binding.

Nickel column affinity chromatography works by passing the soluble cell lysate over a column packed with Ni^{2+} activated resin (**1**). The POI contains a His tag which is a string of 6 histidine residues at the N-terminus of the protein. These histidine residues interact with Ni^{2+} and bind to the column (**2**). The process of this interaction is between the N of two histidine residues interacting with the Ni^{2+} charged nickel. The other soluble proteins are washed through with repeated buffer column volumes (**3**) and the POI is then eluted by addition of imidazole which competes for binding to the Ni^{2+} (**4**) (Figure 6-2).

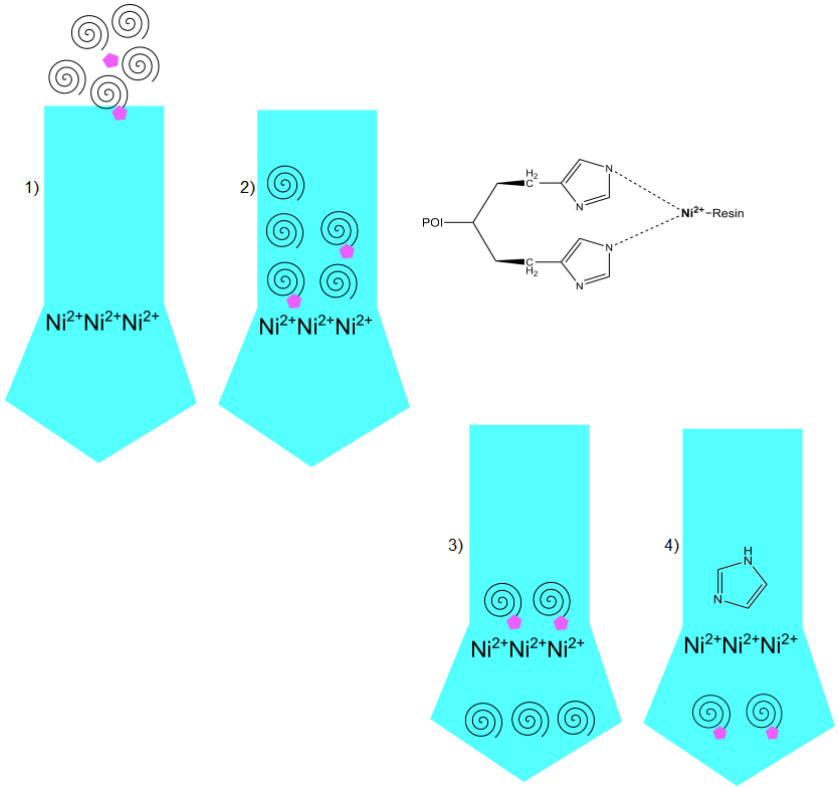


Figure 6-2 Purification of POI containing a His tag by Nickel column chromatrography. Protein mixture is passed over a column packed with activated nickel resin (1-2), the POI His tag attaches to the nickel (3) and is eluted off by competitive Imidazole binding (4).

6-3 Expression of FBA1

The FBA1 sequence is shown below with the putative carbamate site highlighted in red (Figure 6-3).

MASSTATMLKASPVKSDWVKGQSLLLRQPSSVSAIRSHVAPSALTVRAAS
AYADELVKTAKTIASPGHGIMAMDESNATCGKRLASIGLENTEANRQAYR
TLLVSAPGLGQYISGAILFEETLYQSTTDGKKMVDVLVEQNIVPGIKVDK
GLVPLVGSYDESWCQGLDGLASRTAAYYQQGARFAKWRTVVSIPNGPSAL
AVKEAAWGLARYAAISQDSGLVPIVEPEIMLDGEHGIDRTYDVAEKVWAE
VFFYLAQNNVMFEGILLKPSMVTPGAEATDRATPEQVASYTLKLLRNRIP
PAVPGIMFLSGGQSELEATLNLNAMNQAPNPWHVSFSYARALQNTCLKTW
GGKEENVKAAQDILLARAKANSLAQLGKYTGEGESEEAKEGMFVKGYTY

Figure 6-3 FBA1 protein sequence. Containing the end of the plastid coding region (A) and putative carbamate binding site (K).

6-3.1 Test expression

We examined multiple *E. coli* strains for the expression of recombinant FBA1 to determine the optimal strain for protein expression. The FBA1 DNA sequence encoding the complete FBA1 open reading frame (Figure 6.3) within a pET14b vector was purchased from Genscript and transformed into Rosetta, Rosetta 2 and Rosetta Tuner *E. coil* strains (2-3.11). Rosetta is a strain of BL21 cells designed to enhance the expression of eukaryotic proteins that contain codons rare in *E. coli*. These were used for a small-scale test protein expression (2-3.12). The three *E. coli* strains were compared both before and after a 3 h induction with IPTG (1 mM). The protein lysate from three test protein expressions were separated by electrophoresis on an SDS-PAGE gel to determine whether recombinant protein production occurred. The FBA1 protein was identified by its MW of 43 kDa in all lanes (Figure 6-4).

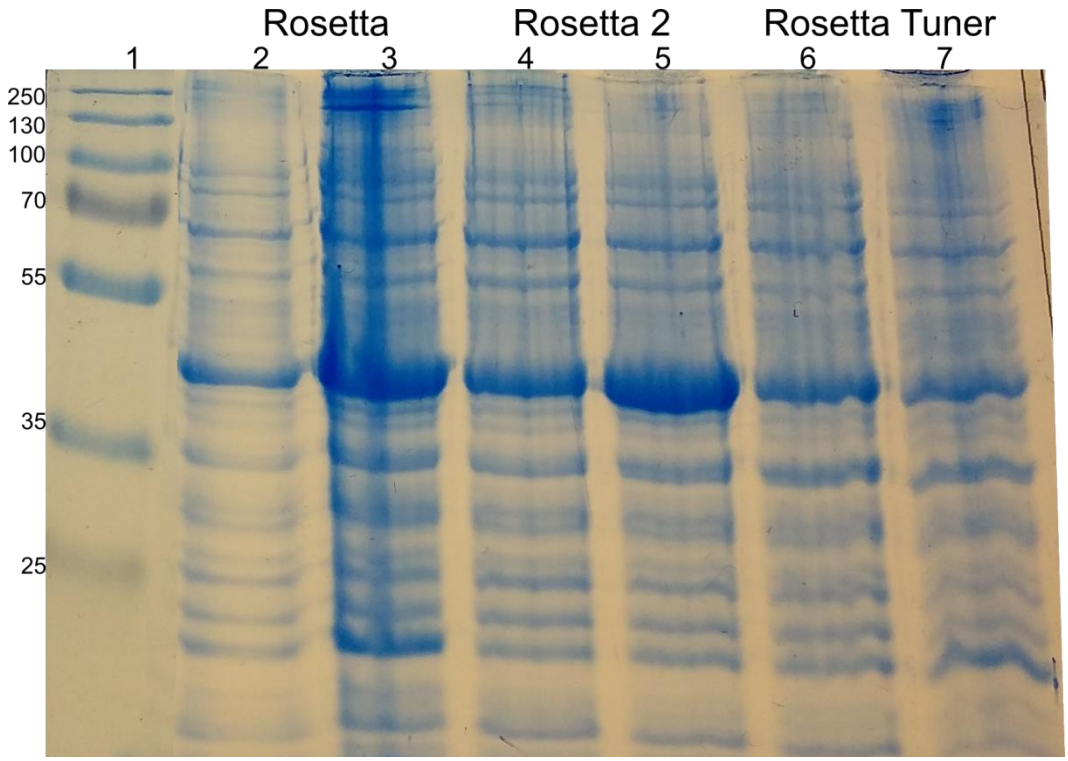


Figure 6-4 SDS-PAGE of test expression of FBA1 using 3 *E. coli* strains Rosetta, Rosetta 2 and Rosetta Tuner. Lane 1 contains a MW marker (kDa), lanes 2, 4 and 6 are pre induction and lanes 3, 5 and 7 are post-induction.

The level of POI production from each cell strain was compared by analysis with SDS-PAGE. A sample from each cell strain pre-induction with IPTG (lanes 2, 4 and 6) and post-induction (lanes 3, 5 and 7) were assessed. All samples contained a band relating to the correct MW for FBA1 (43 kDa) so it was concluded that all cell strains had leaky expression of the recombinant protein. The Rosetta and Rosetta 2 cells were shown to have produced the most protein post-induction so were chosen to investigate further.

To investigate the amount of induced protein that is in the soluble fraction the bacterial pellet produced from the test expression was sonicated in lysis buffer and re-centrifuged to separate the soluble and insoluble fractions. Both the soluble and insoluble fractions of this sample were then separated by electrophoresis by SDS-PAGE for comparison (Figure 6-5).

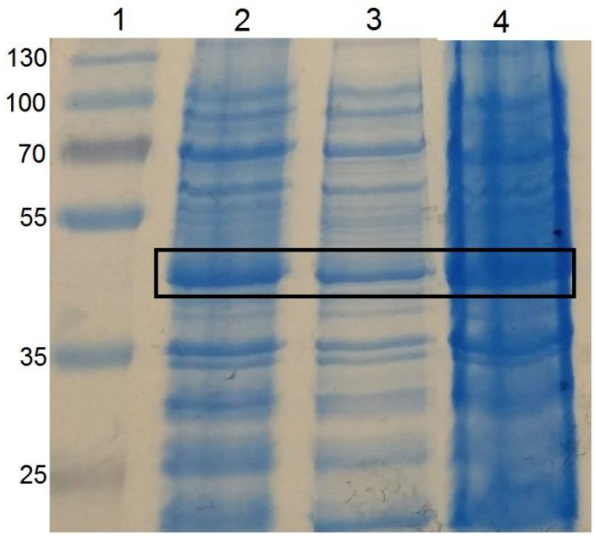


Figure 6-5 SDS-PAGE of Rosetta strain pre-induction (lane 2), soluble fraction post 3 h induction (lane 3) and insoluble fraction post 3 h induction (lane 4) against a MW marker (kDa) (lane 1).

The pre-induction sample (lane 2) is compared to the post-induction soluble protein (lane 3) and insoluble protein (lane 4). The POI was identified by its MW of 43 kDa and is confirmed by the increased band density in the post-induction insoluble sample (lane 4). Though all lanes appear to contain the POI the SDS-PAGE results suggest that the majority of the protein is in the insoluble fraction. This demonstrates the formation of inclusion bodies, which are caused by protein aggregating due to misfolding during protein production (261).

Previous work producing *Arabidopsis* fructose 1,6 bisphosphate aldolase proteins used a DNA sequence which omitted the plastid targeting signal (Figure 6-3 the sequence up to A) (254). It was predicted that the removal of this region would increase the solubilisation of the recombinant protein. Therefore, a new recombinant plasmid in which amino acids 1-48 encoding for the plastid targeting sequence were deleted was produced for further investigation.

This new recombinant plasmid was prepared by carrying out PCR on the desired fragment (sense oligo ggccatatgGCgagcgcgtacgcggacg, antisense oligo gctagttattGCtcagcgg) and confirming the newly reduced DNA sequence size by agarose gel electrophoresis. This new DNA encoding truncated FBA1 was cloned into a pJET1.2 vector system before being digested with BamH1 and Nde1 to produce the correct cohesive DNA ends for ligation into the pET-14b vector.

The newly made recombinant DNA plasmid did not contain any mutations and aligned correctly with the original DNA. This truncated FBA1 (FBA1-WT-Trunc) in vector pET14b was transformed into an *E. coli* BL21 strain and used for the large scale expression of this protein for CO_2 trapping experiments.

6-3.2 Large scale protein expression

The results from the test expression showed the presence of the POI in the non-induced sample suggesting leaky expression in Rosetta cells. This could contribute to the formation of inclusion bodies, so for the large scale expression *E. coli* BL21 pLysS cells were used. The presence of a pLysS plasmid gives tighter expression as it encodes for T7 lysozyme which inhibits T7 RNA polymerase reducing transcription until the *E. coli* cells are induced with IPTG (262).

A large scale expression (12 L) was carried out to produce sufficient protein for purification and CO_2 trapping. As well as changing the *E. coli* strain other protocol modifications were made to reduce inclusion body formation. Inclusion bodies are often formed by protein production occurring too quickly therefore a lower induction temperature of 20 °C and a lower concentration of 20 µM IPTG were used to slow protein production and encourage correct folding (263). Due to this lower temperature a longer induction time of 24 h was used. The results from the expression can be seen in Figure 6-6.

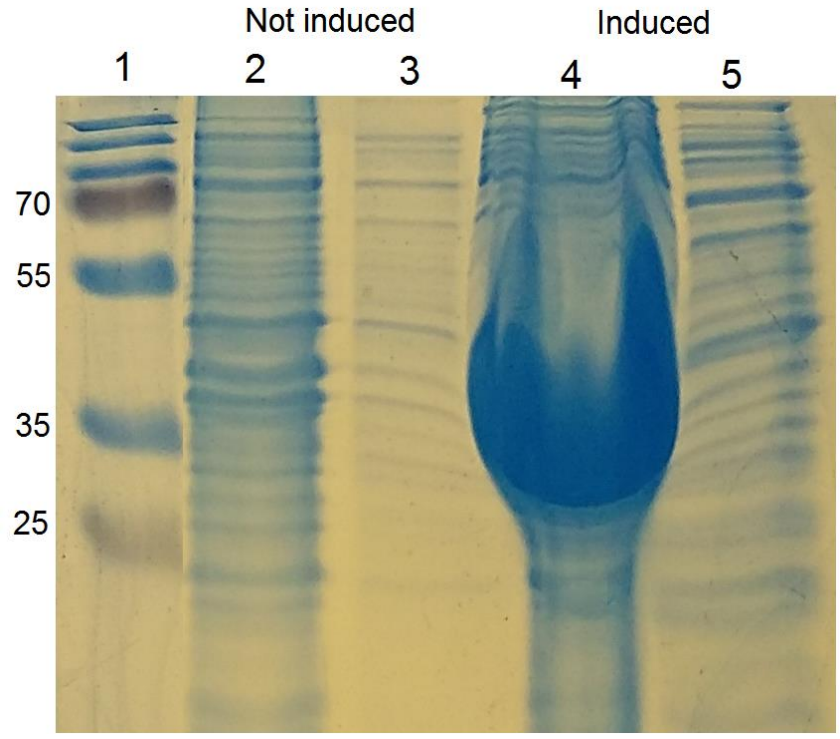


Figure 6-6 SDS-PAGE of BL21 strain test comparing a not induced vs an induced culture. Comparison of insoluble samples (lanes 2 and 4) and soluble samples (lanes 3 and 5).

Samples from an IPTG uninduced and induced culture were compared by SDS-PAGE. From these two samples the insoluble (lanes 2 and 4) and soluble (lanes 3 and 5) proteins were evaluated. Lane 4 contains the insoluble protein from the induced culture and contains a very large protein band corresponding to FBA1-WT-Trunc inclusion body formation. This band is identified as the FBA1-WT-Trunc as it corresponds to the newly truncated MW of 38 kDa and can clearly be seen to be increasing in the uninduced culture. Protein purified from inclusion bodies was refolded under buffer conditions to encourage the correct structural formation of the native protein.

6-3.3 Inclusion bodies

Inclusion bodies are formed of insoluble protein due to misfolding. Inclusion bodies can be refolded to encourage the correct soluble formation of the protein structure and was carried out using a QuickFold refolding kit (AthenaES) (chapter 2; 2-3.23). Samples of the aggregated FBA1 (100 µL) were dialysed into a range of buffers overnight. Several buffers produced solubilised protein but the optimal buffer for FBA1 refolding was 50 mM Tris-HCl pH 8.5, 240 mM NaCl, 10 mM KCl, 1 mM EDTA, 0.5% triton X-100, 1 mM DTT (Figure 6-7).

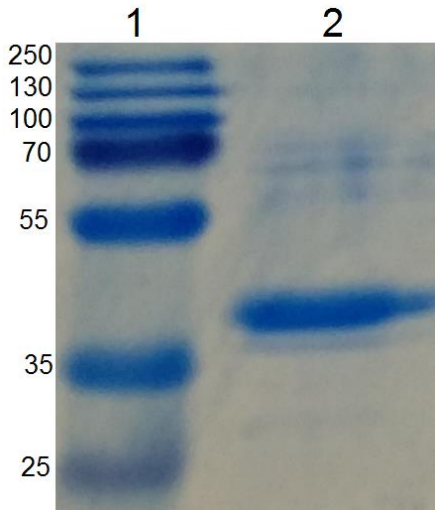


Figure 6-7 SDS-PAGE of refolding of FBA1-WT from inclusion bodies showing the MW marker (lane 1) and a large soluble FBA1-WT protein band from the the best refolding condition (lane 2).

The supernatant from the refolding was examined by SDS-PAGE with a protein corresponding to FBA1 observed at 38 kDa (lane 2) for the optimal refolding buffer. This re-folded protein was then tested for its aldolase activity via assay (6-5) compared to a purchased aldolase control protein (Sigma Aldrich- A2714). No aldolase activity was seen for the refolded FBA1-WT-Trunc.

Refolding can only show that the protein is now in a solubilised state not that it is correctly folded to form the active site. As this has not produced active protein this refolded inclusion body protein was not used to confirm carbamate formation. Instead the purification of the soluble protein produced during expression was further developed.

6-3.4 Improved soluble protein conditions

Some soluble protein at 38 kDa had been observed in the test expression (Figure 6-6 lane 4) previously. As the refolded insoluble protein was shown to be inactive, the activity of the SN of an FBA1-WT-Trunc culture was measured to examine whether any soluble protein production had occurred (Figure 6-14). Aldolase activity was identified within the sample so further development was directed toward production of soluble protein. Other methods to help solubilise aggregating recombinant proteins are by increasing the concentration of osmolytes or by addition of chaperone proteins. In this work the concentration of osmolytes was altered to improve amount of protein remaining in the soluble fraction. The method used here was the addition of a final concentration of 2 mM betaine in the *E. coli* growth medium (264). Betaine is one of the best known stabilising osmolytes; it helps the cell adapt to osmotic stress and assists in stabilising folded protein conformations (265). This additive has been used previously for recombinant protein expression within an *E. coli* system (266) so was adapted here to aid the reduction in inclusion body formation. Samples of the protein expression were taken after three and 24 h post induction and the soluble protein electrophoresed on an SDS-PAGE gel (Figure 6-8).

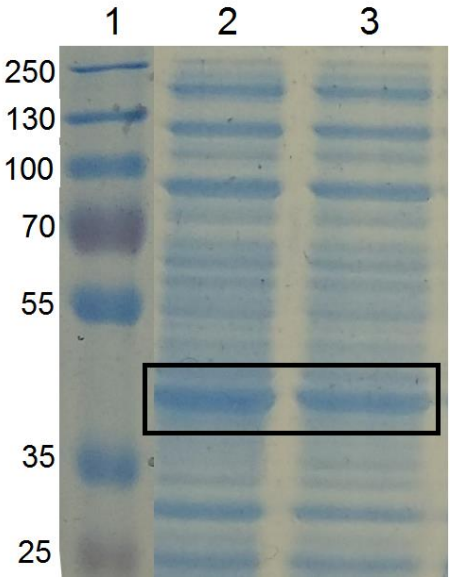


Figure 6-8 SDS-PAGE of test conditions to improve amount of soluble FBA1-WT, showing MW (kDa) marker (lane 1) and no increase in the amount of protein in the soluble layer after 3 h (lane 2) and 24 h (lane 3).

The amount of protein observed in the soluble fraction of the *E coli* lysate was seen to increase with the addition of betaine and a lower induction temperature of 17 °C and the use of 24 h induction period (lane 3) so this was the method used to express FBA1-WT-Trunc for assay and trapping.

6-3.5 Soluble FBA1-WT-Trunc purification

Affinity column purification was necessary after the large scale growth to purify the desired protein from the remainder of the *E. coli* soluble proteins. The FBA1-WT-Trunc protein contains a His tag on the N-terminus so purification was carried out using a Ni^{2+} column. The His tag binds to the Ni^{2+} attaching it to the column resin, the remaining *E. coli* proteins are washed through and the POI is then eluted with increasing concentrations of imidazole to produce purified FBA1-WT-Trunc (Figure 6-2).

The initial column chromatography purifications produced elution fractions still containing many other proteins resulting in an impure FBA1-WT-Trunc sample (Figure 6-9).

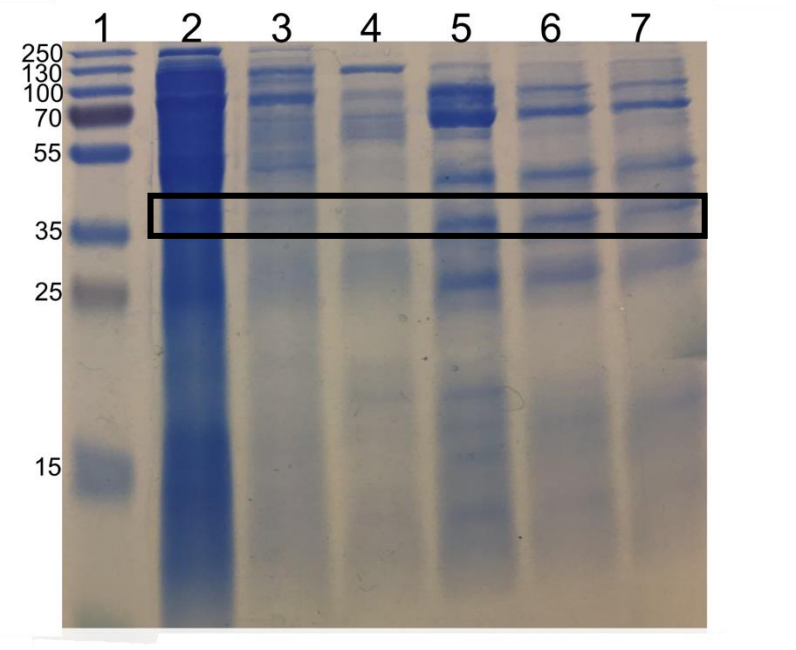


Figure 6-9 SDS-PAGE of purification of FBA1-WT-Trunc using a Ni^{2+} column showing the MW marker (kDa) (lane 1), flow through (lane 2), wash fractions (lanes 3 and 4) and the elution fractions (lanes 5-7) with the POI FBA1-WT identified by the correct weight and highlighted with a black box.

The elution fractions (lanes 5, 6 and 7) all contain the POI as well as several other contaminating proteins. The wash steps (lanes 3 and 4) do not contain the POI and there is insufficient washing occurring to remove all of the contaminating proteins.

As the POI appeared to be present in the elution fractions, but contaminated with other *E coli* proteins, the purification procedure was modified. A new protocol increased the incubation time of the sample on the resin for 30 min to 2 h to encourage the POI to bind. The new protocol also increased the volume of wash buffer passed over the column prior to elution of the POI. This volume was increased from 3 column volumes to 20 to remove the contaminating proteins. These steps greatly improved the purity of the eluted protein (Figure 6-10).

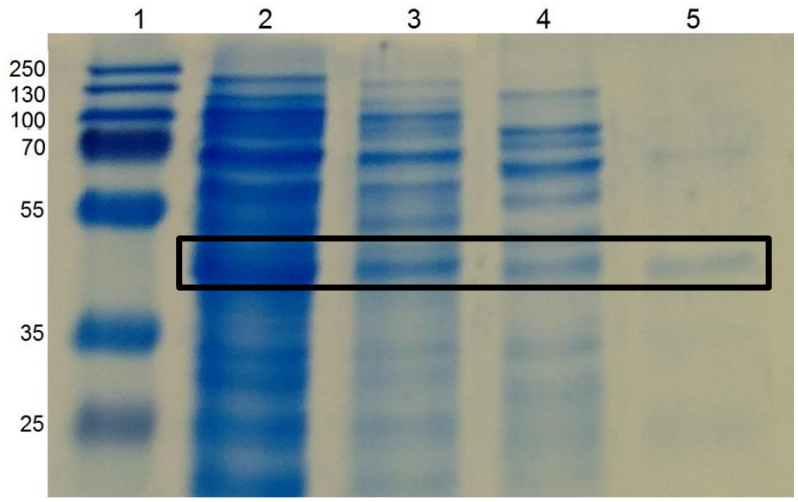


Figure 6-10 SDS-PAGE of improved purification of FBA1-WT-Trunc with increased volume of wash buffer, showing MW marker (kDa) (lane 1), flow-through (lane 2), wash fractions (lanes 3-4) and elution (lane 5).

The SDS-PAGE contained a MW marker (lane 1), the cell lysate flow through the column (lane 2), two wash fractions (lanes 3 and 4) and an elution fraction (lane 5). All lanes contain FBA1-WT-Trunc protein (highlighted by black box) but the elution fraction in lane 5 shows a very pure sample, this purified protein was then used for carbamate trapping and assayed for protein activity (6-4 and 6-5.2).

6-3.6 Construction of mutant FBA1 protein

The hypothesised carbamate site of K293 in FBA1 (Chapter 5, Figure 5-16) was mutated from a lysine to an alanine, which is not capable of forming a carbamate, to investigate the importance of this amino acid in aldolase activity in response to CO_2.

This mutation was carried out using PCR mutagenesis with the sense oligo gttgcgagctacaccctgGCgctgctgcgtaaccgtatc and the antisense oligo: gatacggttacgcagcagcGCcagggtgtagctcgcaa. These primers were used to amplify the whole FBA1-WT-Trunc DNA plasmid by PCR (2-3.23). This template DNA plasmid was then digested using the restriction enzyme Dpn1 which digests the methylated template DNA (267) before transformation of DH5α with the reaction mixture.

This mutated plasmid DNA was sequenced and aligned with the truncated aldolase construct to confirm the presence of the single point DNA mutation at lysine 293 converting it to alanine (FBA1-K293A).

The K293A mutated plasmid was transformed into *E. coli* cells BL21 strain and grown on a large scale under the conditions developed for the wild-type FBA1 to produce FBA1-K293A-Trunc. However the process of producing this mutant was never fully purified due to time constraints.

6-4 Purified FBA1 carbamate trapping

Once FBA1-WT-Trunc protein had been expressed and purified the CO_2 trapping experiment was carried out using the conditions developed in chapter 4. An additional trapping experiment with a lower concentration of sodium bicarbonate of 1 mM which at pH 7.4 is equivalent to 73.5 µM CO_2 which approximates hypothesised intracellular concentration within plants (268).

These trapping experiments were carried out to validate the carbamate identified through the proteomics screen (chapter 5) and to demonstrate that the carbamate would also form under lower [CO_2] applicable to a physiological leaf environment. The CO_2 trapping at both concentrations identified the carbamate modification at K293 (Figure 6-11).

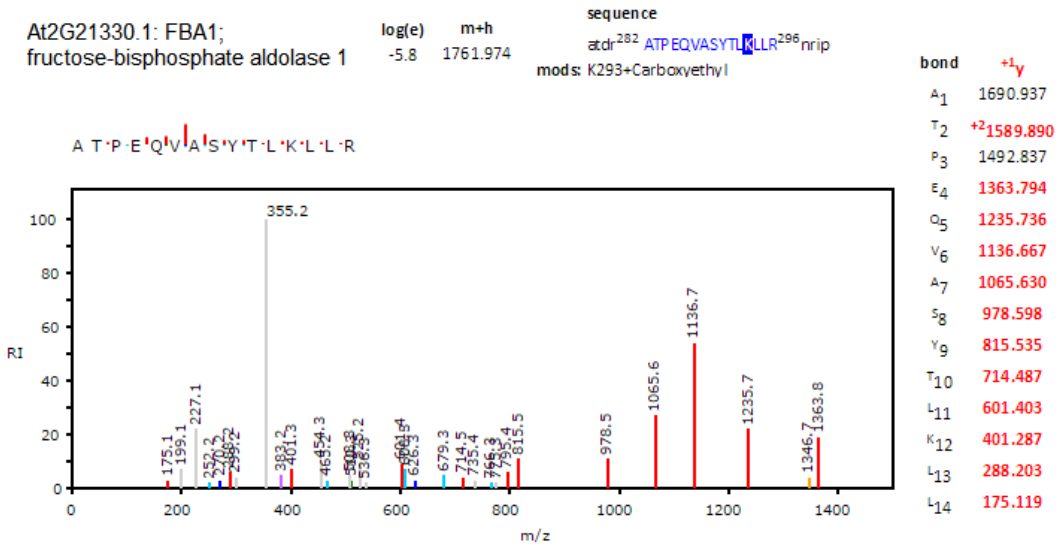


Figure 6-11 Image of fragmentation pattern of the FBA1-WT-Trunc peptide containing the K293 carbamate binding site displayed from Tandem.

The trapped carbamate was visualised using the Tandem software at the putative site (K293). This hit was found with a high confidence score of −5.8 and almost a complete peptide fragmentation pattern. These results demonstrate a high level of confidence for the hypothesised carbamate site. The success of this trapping experiment validates the previous proteomics carried out in chapter 5 lending support to the other possible sites of carbamylation identified on other proteins.

6-5 Fructose Bisphosphate aldolase 1 assay

6-5.1 Aldolase reaction mechanism

There are two reactions catalysed by an aldolase enzyme, the cleavage and the condensation reaction. Both reactions play a role in the sugar metabolic pathway of all organisms (269). The cleavage reaction in glycolysis is the breakdown of fructose-1,6-bisphosphate (F-1,6-BP) into glyceraldehyde-3-phosphate (G3P) and dihydroxyacetone phosphate (DHAP) (270). The reaction mechanism of this cleavage is shown in Figure 6-12.

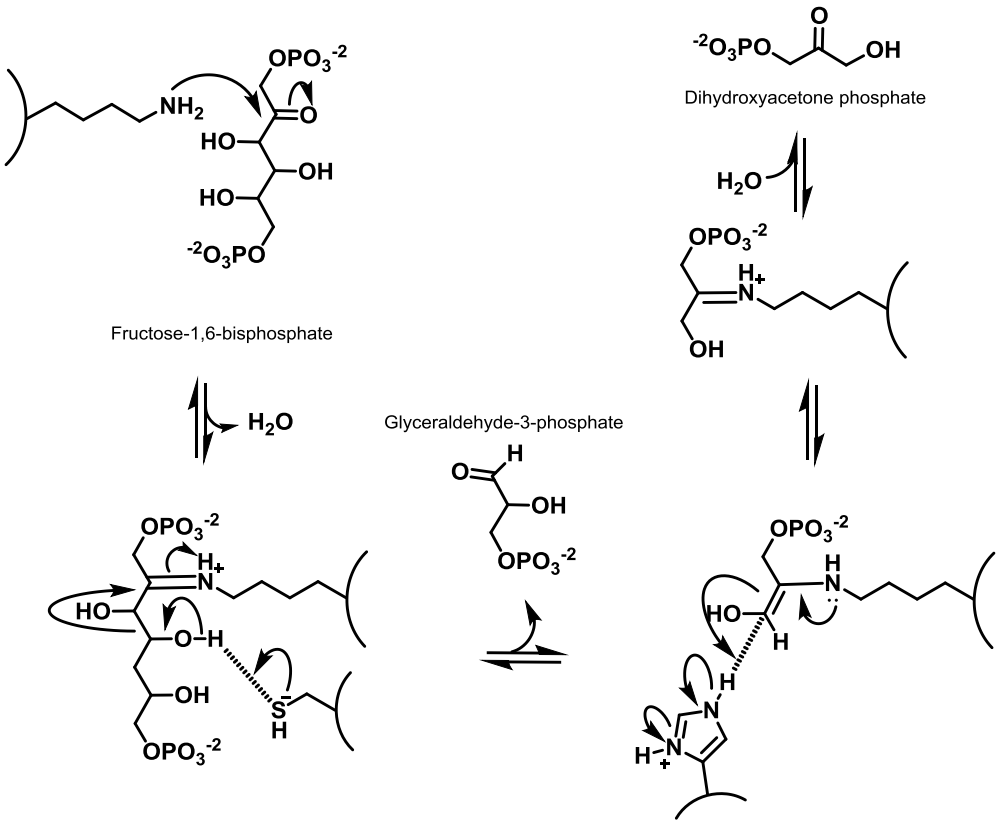


Figure 6-12 Steps in the binding of F-1,6-BP to the aldolase enzyme AS and it's breakdown to G-3-P and DHAP.

The F-1,6-BP substrate interacts with a lysine and cysteine amino acid to release G-3-P and then the remaining chain interacts with a histidine residue and is rearranged to form DHAP which is then released.

In order to use this reaction as an assay it was coupled to the oxidation of reduced nicotinamide adenine dinucleotide (NADH) to nicotinamide adenine dinucleotide (NAD^+) by α-glycerophosphate dehydrogenase (GDH) and triosephosphate isomerase (TPI) enzymes (Figure 6-13).

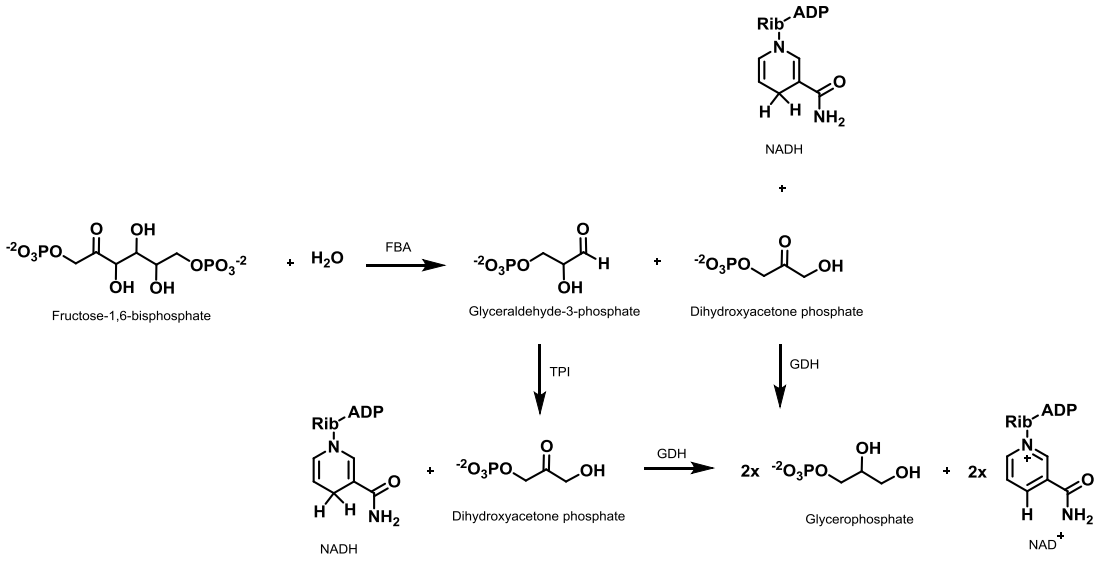


Figure 6-13 Mechanism of coupling the aldolase cleavage reaction to the oxidation of NADH to form an activity assay to investigate enzyme activity of FBA1.

6-5.2 Synthesised FBA1-WT and FBA1-K293A activity assay

The aldolase assay was used to assess the activity of the synthesised FBA1-WT-Trunc and FBA1-K293A-Trunc to confirm that the CO_2 trapping had been performed on a functionally active enzyme. An aldolase assay was carried out on unpurified SN of FBA1-WT-Trunc and FBA1-K293A growth cultures to investigate whether any active aldolase protein is being produced. The results of this assay showed small amounts of activity from both supernatant cultures compared to a control sample without any aldolase enzyme (results not shown). The most likely reason for such small activity levels is due to the small concentration of FBA protein within the unpurified SN.

The assay was then repeated using purified FBA1-WT-Trunc protein shown in Figure 6-10, this was to investigate the activity of the protein post-purification (Figure

6-14). This assay was conducted to investigate the presence of active protein being produced during expression within the *E. coli* growth culture.

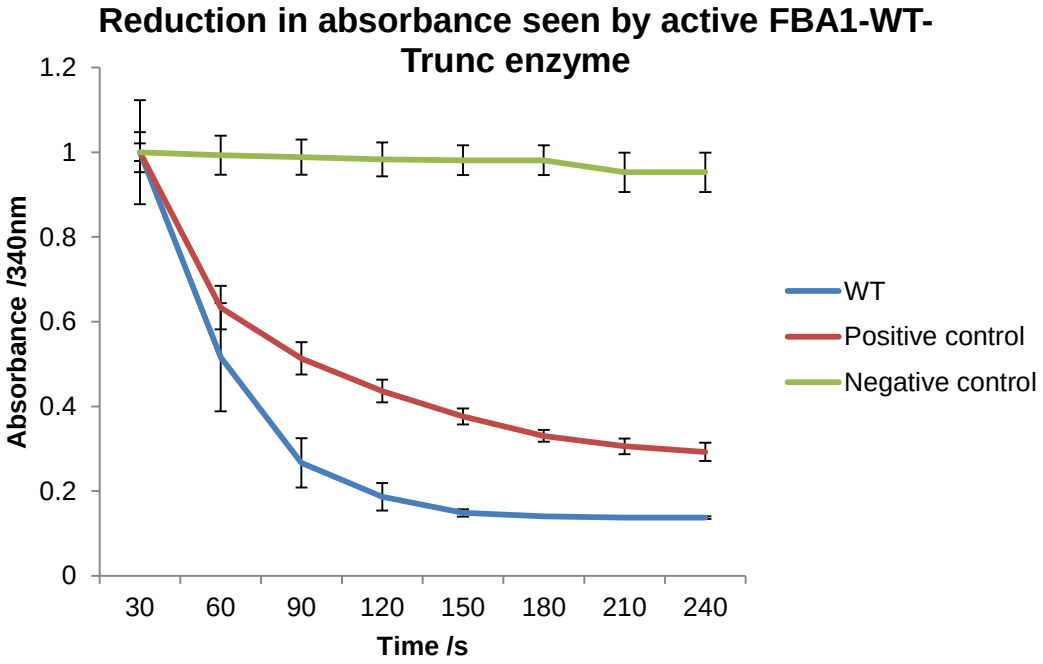


Figure 6-14 Graph to show the aldolase activity of FBA1-WT-Trunc compared to a control sample not containing any enzyme.

Figure 6-14 shows a reduction in absorbance in the presence of purified FBA1 that correlates to the oxidation of NADH (blue line) compared to a no-enzyme negative control (green line) and a positive control of purchased Aldolase (SigmaAldrich) (red line) over a three-minute time scale. This indicates that the soluble protein is in the correct conformation to produce an active site capable of cleaving the F-1,6-BP substrate. This result demonstrates that the soluble protein purified has aldolase activity after purification. This validates that the trapping experiments were undertaken with active protein.

6-5.3 CO_2 dependence on activity

After the assay of FBA1-WT-Trunc demonstrated active protein the next step in the line of investigation was to investigate the effect of CO_2 on this activity. One of the largest challenges in this investigation was the complete removal of CO_2 from the assay due to the presence of CO_2 in the atmosphere dissolving within the reaction solution.

The usual methods of removing carbon dioxide from a solution are to use a carbon dioxide scrubber, sonication or displacement by sparging with another gas through the solution. The scrubber method does not completely remove the carbon dioxide so would not be effective here and sonication cannot be used on a solution containing proteins as this will denature the protein and remove activity. Therefore sparging with argon (Ar) was used on all of the solvents and the mixture was incubated under an Ar atmosphere prior to substrate addition.

In order to investigate the effect CO_2 has on the Km or the Vmax the investigation was carried out using two different concentrations of substrate (fructose-1,6-bisphosphate). A low (40 µM) and highly saturated (400 mM) substrate concentration (253). The use of 400 mM substrate concentration will demonstrate the maximum rate for the enzyme due to the saturation of the enzyme active sites. A low concentration of substrate will investigate the Km which is the affinity of the enzyme for the substrate as the substrate is now limiting. To thoroughly investigate the Vmax and Km values of an enzyme a range of substrate concentrations need to be investigated and plotted.

Unfortunately the preliminary results (not shown) from the first experiment showed no significant difference between the Km and Vmax and repeat experiments were not undertaken.

The FBA1-K293-Trunc mutant protein will also provide information about the effect of CO_2 interaction within the assay however; this protein was not successfully purified within the time scale of this study.

6-5.4 Protein structural information

Protein fold searches using the Phyre^2 protein homology/analogy recognition engine version 2.0 (271) were undertaken using the open reading frame of FBA1, using both normal and intensive modelling modes. The model was based on the crystal structure of the rabbit fructose diphosphate aldolase (PDB accession number 1FDJ; 100% coverage, 48% identity) (272) and sequence alignments were generated by the Phyre^2 server. Side chain packing and energy minimization was performed using GalaxyRefine (273). Figure was generated using the PyMOL molecular graphics system (274) with assistance from Phil Townsend (Figure 6-15).

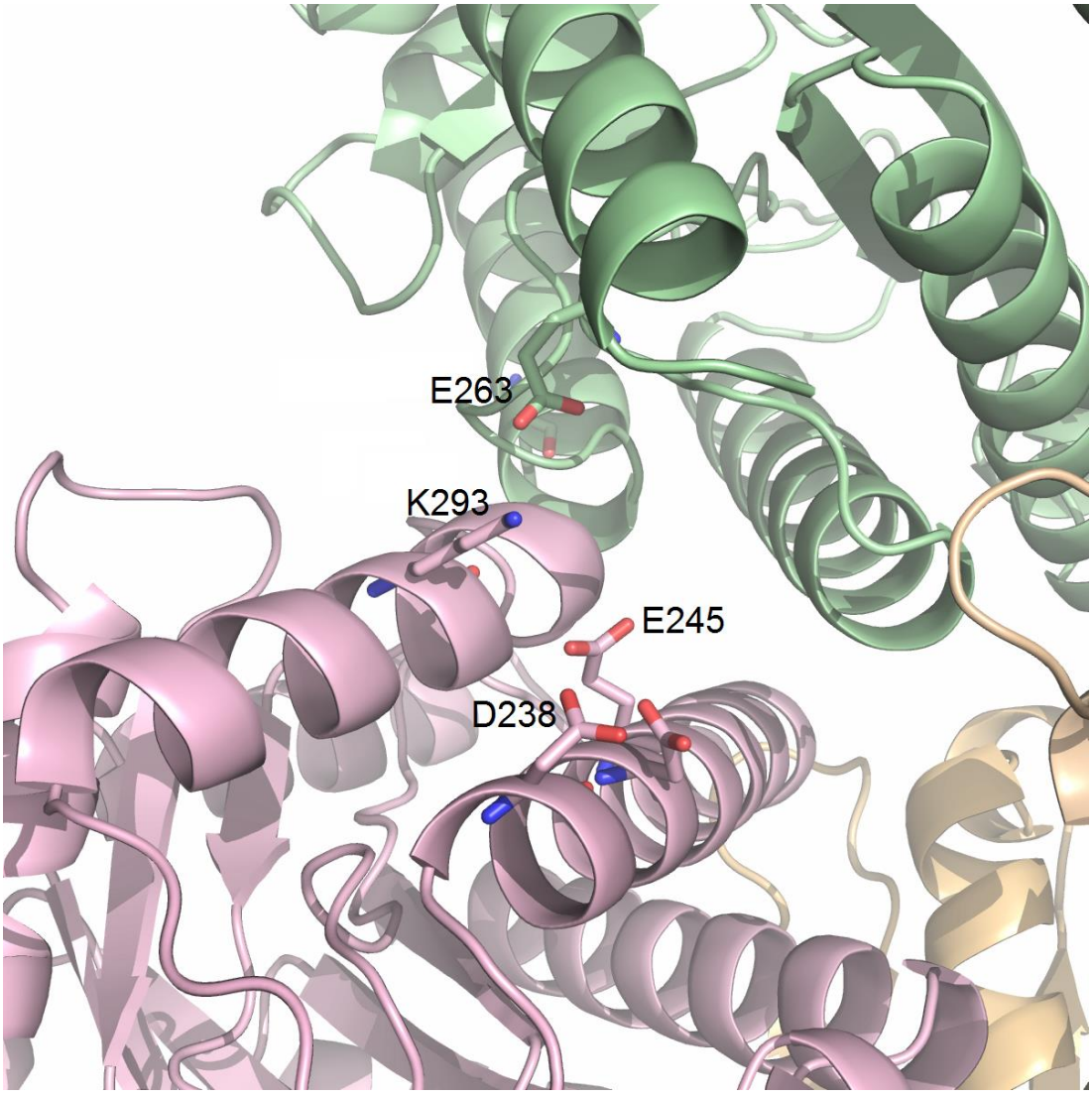


Figure 6-15 Model based on crystal structure of the rabbit fructose diphosphate aldolase – PDB accession number 1FDJ showing the residues surrounding the K293 putative carbamate site.

The residues surrounding the K293 putative carbamate binding site are all residues with a carboxylic acid side chain (E263, D238 and E245).

The production of a carbamate at this location causes the conversion of the lysine residue from a neutral to acidic charge state. Based on previously identified carbamate sites it is likely that the new charge could contribute to the surrounding negative residues possibly involved in stabilisation, maybe for a coordinated metal binding site. It could also be involved in encouraging a structural change in folding of the protein due to the new negative charge causing repulsion with the surrounding negative residues. Future crystallisation of FBA1 with and without CO_2 could help distinguish between these hypothesised roles.

6-6 Conclusion

This chapter has described the process of validation of a newly discovered carbamate site within FBA1, a protein previously unknown to bind CO_2.

The results discussed in this chapter have shown that the carbamate formation occurs on purified FBA1 at the same site as previously identified within Chapter 5. This demonstrates confidence that this formation is a true carbamate and not through lysate interactions. This carbamate has also been confirmed at lower concentrations of CO_2 more accurately representing a plant cellular environment and demonstrating the ability of the trapping reaction to still be accurate at this scale.

The purified protein was assayed and established to be in its active form. This further encourages the belief that the carbamate found in the proteome screen is a true site and that the method for extracting the protein lysate is not causing inactivity of protein prior to the trapping experiment.

This work adds value to the other hits on unknown proteins that were discovered within the leaf lysate screen described in Chapter 5 and demonstrates that there are many new sites of CO_2 interaction that were previously unknown.

6-7 Future work

The method of protein production and purification needs further development to reduce the production of inclusion bodies and therefore produce a larger scale of soluble active protein. In some previous cases tags have been used to increase the solubility of recombinant proteins, for example the Fh8 fusion tag (275), which could be investigated.

Due to time constraints only preliminary work on the CO_2 dependence of the recombinant wild type FBA1 was performed. These results show a very promising effect of CO_2 on the activity of FBA1 but further investigation is necessary for conclusions.

The assay dependence on CO_2 could be improved by further replicates so that statistical analysis to show a significant difference found in the presence of CO_2 can be assessed. Further validation to the experimental relevance of CO_2 within the reaction could be investigated with the use of the K293 mutant to observe the effect seen when a carbamate cannot form at the confirmed site.

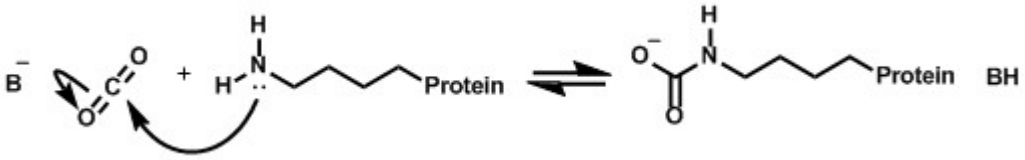
As well as the work discussed here further work of validation needs to be carried out on more of the new carbamate ‘hits’ found in chapter 5 to confirm their sites and mutations made to investigate the role of the carbamate within the leaf system.

Chapter 7: Synopsis

7-1 Introduction

Despite the abundance of CO_2 within cellular systems, little is known about its molecular interactions with protein. Further information about the molecular CO_2 interface within proteins could provide essential information for numerous areas of research.

A carbamate is the nucleophilic attack of an uncharged amine on CO_2 (Scheme 7-1) and can occur under cellular conditions of CO_2 concentration and pH.



Scheme 7-1 Nucleophilic attack of an uncharged amine on carbon dioxide to form a carbamate.

Within proteins known to form carbamates the carbamylated amino acid functions in regulatory or catalytic roles (e.g. β-lactamase and RuBisCO). This highlights the possible importance the carbamate post-translational modification can have on a protein system (84).

The work in this thesis describes the development of a novel method for trapping carbamates on protein under physiologically relevant conditions. This chapter summarises the results outlined in this investigation as well as highlighting the important benefits of studying the identification of carbamylation sites on protein.

7-2 Partial synthesis of a carbamate trapping reagent

Chapter 3 of this thesis describes the process for design of a synthetic TMS-DAM derivative to trap carbamates by the transfer of a methyl group. This synthesis was undertaken in parallel with an investigation into the use of Meerwein's reagent for carbamate trapping (chapter 4). Due to the success seen with Meerwein's reagent the synthesis of the TMS-DAM derivative was not taken to completion, **1** (Figure 7-1). (Although the achieved synthetic steps were well developed with high yields.)

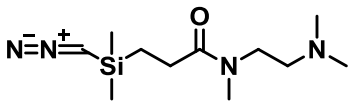
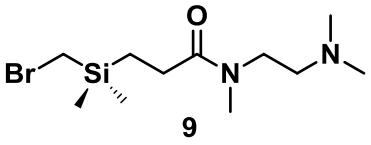


Figure 7-1 Diazo trapping molecule 1.

The synthetic work covered here was modified from several previously published procedures to synthesise TMS-DAM (141,156,160,276). The modifications were to include an amine group into the synthetic derivative to introduce water solubility into the molecule. Previous work with TMS-DAM involved the use of mostly organic solvents which is a clear barrier to its use under physiologically relevant conditions (150). The final synthetic product of Chapter 3 can be seen in Figure 7-2 ([(bromomethyl)(dimethyl)silyl] [(methylamino)3-(dimethylamino)ethyl] Propamide **9**)



9

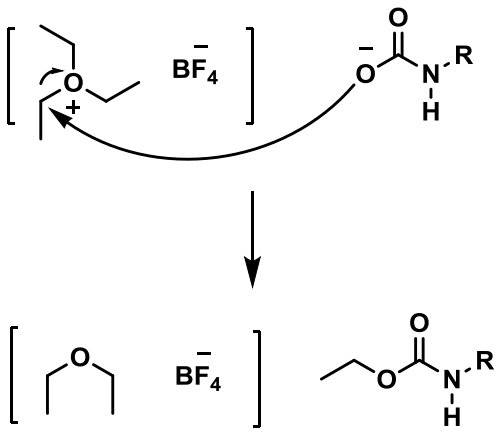
Figure 7-2 Synthetic final product from Chapter 3.

Though this molecule is not capable of carbamate trapping it has been suggested to have an effect as a protein kinase inhibitor due to being a small water-soluble molecule with an easily altered halogen group. This could be tested using a simple binding affinity assay (277).

7-3 Development of the trapping methodology using Meerwein's reagents

Carbamylation has not been fully investigated as a post-translational modification due to its labile nature (83-85). In addition to its labile nature, its ready reversibility has meant that carbamates have previously been considered unlikely to be of biological significance (83). Until recently (97) this has meant that protein carbamylation has not elicited significant interest since their first mention in the early 1900's (82).

The carbamate trapping method developed in this investigation is described in detail in Chapter 4. This method chemically traps CO_2 on proteins using the reagent TEO which transfers an ethyl group to the carbamate (Scheme 7-2) creating a modification robust enough for downstream analysis by MS.



Scheme 7-2 Process of ethyl transfer from TEO to a carbamate.

This research has produced a method capable of removing the labile nature of carbamates and thereby completely transforming the field of carbamate study.

7-4 Newly discovered carbamate sites

This novel method was applied to a soluble protein lysate extracted from *Arabidopsis* leaves to screen for the presence of proteins modified by carbamylation (Chapter 5). These complex protein mixtures were digested with trypsin and fractionated by cation exchange before being analysed by ESI-MS. This work revealed new carbamate modifications at previously unknown sites. Though all the newly discovered proteins have different cellular functions they are all functions involving CO_2 (e.g. photosynthesis and guard cell function).

Investigating the effect of CO_2 on plant proteins with functional roles in photosynthesis links to discussion of the effect of rising atmospheric CO_2 levels on crop growth (134). Multiple investigations into the effects of rising CO_2 on crops have not agreed upon what result this will have on crop yield (61,132). These debates surround the lack of a sufficiently well-defined environment to perform these studies, the inadequate data so far collected, and the limited time frame these studies have been carried out over (62). If more were known about the proteins directly regulated by CO_2 then downstream physiological effects would be much easier to predict. Investigating CO_2 binding proteins within a system will provide crucial information about the regulation of CO_2 on cellular processes.

Though the work described here focused on a plant leaf lysate, the developed methodology is not limited to use in plants but could be used to provide novel insight into how any cellular system is affected by interactions with CO_2. For example NF-κβ has been shown to link CO_2 to immunity and inflammation in mammalian cells by the rapid movement of IKKα to the nucleus in response to elevated CO_2. Cummins *et. al.* 2010 suggests that this implies a molecular CO_2 sensor in mammalian cells. The ability to screen the mammalian proteome for sites of carbamate formation could determine the mechanism of the effect of CO_2 on inflammation.

It has previously been shown that increased levels of CO_2 reduce growth rates of bacteria (278). This can be a problem for the commercial production of recombinant proteins. Further knowledge of the affected pathways could reduce this problem by rationale engineering of protein carbamates to enhance yield (279).

Further information about CO_2 interactions could also be of significance to clinical situations involving elevated P_{CO2}. For example, there is a debate (280,281) concerning the influence of permissive hypercapnia on patients with acute respiratory distress syndrome (ARDS). Permissive hypercapnia is the elevation of arterial CO_2 associated with artificial mechanical ventilation designed to minimise lung stretch injury. These respiratory rates are related to tidal volumes, low tidal volumes minimise lung injury but may contribute to hypercapnia. Curley *et. al.* 2013 describes the benefits seen by a reduction in inflammation caused by hypercapnia but Beitler *et. al.* 2013 claim that hypercapnia worsens pulmonary hypertension and that there is insufficient evidence to draw meaningful clinically relevant conclusions. Both discussions note that advances in understanding of the mechanism of increased CO_2 within this system would enable a further understanding of the benefits to ARDS patients. The developed methodology for identifying protein carbamates is a potential route forward.

Previous work on discovering carbamates has occurred through their fortuitous discovery by X-ray crystallography (102,112,282). Crystallography as a route to carbamate discovery has several disadvantages; it is highly work intensive, does not work for all proteins and is performed under artificial conditions. The work here has described a method capable of screening a large complex mixture of proteins to identify the CO_2 binding sites. The CO_2 trapping is achieved under conditions that are physiological with regards to the aqueous environment and pH of the system. It has been shown by the activity of extracted FBA1 (chapter 6) that the method of extracting the soluble leaf lysate is not altering the activity of the proteins prior to trapping.

7-5 Validation of a previously unknown protein carbamylation site

One newly discovered carbamylation site was K293 of FBA1 from *Arabidopsis.* This carbamylation site was further investigated to confirm that the carbamate trapping was occurring on a correctly folded and active protein (Chapter 6). These validation experiments were performed by the production of a purified recombinant protein which was assayed for protein activity.

These experiments demonstrated that the CO_2 trapping was occurring on an active protein and provides supporting evidence that the carbamates discovered through the proteomic screen (chapter 5) represent real CO_2 binding sites and are not experimental artefacts of the complex protein mixture or analytical methods.

7-6 Conclusions

The results described in this thesis demonstrate a novel methodology to trap carbamates on proteins creating a robust modification for downstream MS analysis. This ground breaking technology allows the identification of proteins targeted by CO_2.

The developed methodology functions under aqueous conditions closely matching a physiological environment and in so doing creates the first method of observing carbamates under a more natural cellular setting without the chance of artefact introduction.

The methodology provides a technology to identify sites of carbamate formation and will allow a significant expansion of our fundamental understanding of protein regulation by CO_2. This method can be taken forward to eventually establish the extent to which CO_2 interacts with the proteome, forming a complete 'carbamylome'.

This breakthrough can be applied to all research concerning CO_2 protein interactions. For example research into greener methods of carbon capture by Drummond *et. al.* 2010 looked for insight into new technologies based on RuBisCO due to its ability to sequester 10^{11} tons of CO_2 a year. This binding of CO_2 is through carbamylation.

7-7 Future work

A limitation of the work described here is the number of proteins identified in the proteomics screen. Method development into solubilisation of the post-trapping samples would allow for more proteins to be identified. A description of a more aggressive method of solubilisation called FASP, involving 8 M Urea, has been used in other studies of difficult to solubilise proteins and could be introduced here (283).

Previous proteomic studies have identified thousands of proteins (211,212). This was achieved by thorough pre-fractionation of the sample prior to MS. Subcellular fractionation of the leaf cellular compartments could be introduced with the use of Percoll density centrifugation. Based on the results discovered here this approach could be targeted towards the chloroplast as it is expected this is a location containing many proteins interacting with CO_2.

An increase in the number of proteins identified could also be obtained by improvement of the fractionation of samples post-digest. A new methodology has been decided which will involve fractionation of the sample into 24 fractions by OffGel electrophoresis prior to injection into the ESI-MS machine. This will greatly reduce the complexity of the peptide mixture prior to MS.

Many new sites of carbamate formation were identified in the *Arabidopsis* screen in Chapter 5. Though Chapter 6 was able to ascertain that the extracted soluble protein lysate did contain an active aldolase, future work should validate CO_2 trapping on recombinant proteins corresponding to other targets.

A further important aspect of future validation experiments would focus on the effect of the formation of a carbamate on protein function. The development of methods capable of assaying the protein with complete CO_2 removal from the experiment is necessary to be able to investigate this further. Such methods could then be used to investigate wild type protein and proteins mutated at the carbamylated lysine to identify the biochemical relevance for the carbamate formation. Previously identified carbamylation sites have functions in structural change (Hb (84)), active site stabilisation (Urease (108)) and enzyme catalysis (RuBisCO (83)).

Where structural information is already available for a protein, the identified carbamylation site could be analysed to search for nearby residues and look for possible interactions that the carbamate could be stabilising providing a rationale for the influence of carbamylation on protein function (e.g. Cx26, (97)).

Another avenue for future work is to investigate the physiological significance of the protein carbamylation site identified. Such work could be performed in *Arabidopsis* through the use of *Arabidopsis* knock out plants to investigate the importance of the newly identified carbamylated proteins. Reintroduction of the wild type gene, or a gene mutated at the carbamylation site, could be used to rescue these knock out mutants and investigate the specific function of the carbamate in the physiological response to CO_2 in the whole organism.

The future of this work is heading towards trapping within an *in vivo* system. TEO is a small molecule with no overall charge and is hypothesised to traverse a cellular membrane which provides the possibility of conducting the same methodology upon whole cells.

Chapter 8: Bibliography

1. Joshi, H. M., and Tabita, F. R. (1996) A global two component signal transduction system that integrates the control of photosynthesis, carbon dioxide assimilation, and nitrogen fixation. *PNAS* **93**, 14515-14520
2. Hetherington, A. M., and Raven, J. A. (2005) The biology of carbon dioxide. *Current Biology* **15**, R406-R410
3. Gutknecht, J., Bisson, M., and Toesteson, F. (1977) Diffusion of carbon dioxide through lipid bilayer membranes. Effects of carbonic anhydrase, bicarbonate and unstirred layers. *Journal of General Physiology* **69**, 779-794
4. Boron, W. F., Endeward, V., Gros, G., Musa-Aziz, R., and Pohl, P. (2011) Intrinsic CO_2 Permeability of Cell Membranes and Potential Biological Relevance of CO_2 Channels. *ChemPhysChem* **12**, 1017-1019
5. Smith, K. S., and Ferry, J. G. (2000) Prokaryotic carbonic anhydrases. *FEMS Microbiology Reviews* **24**, 335-366
6. Khalifah, R. (1973) Carbon Dioxide Hydration activity of carbonic anhydrase: Paradoxica consequences of the unusually rapid catalysis. *PNAS* **70**, 1986-1989
7. Lindskog, S. (1997) Structure and mechanism of carbonic anhydrase. *Pharmacology & Therapeutics* **74**, 1-20
8. Frommer, W. B. (2010) CO2mmon Sense. *Science* **327**, 275-276
9. Supuran, C. T., Scozzafava, A., and Casini, A. (2003) Carbonic anhydrase inhibitors. *Medicinal Research Reviews* **23**, 146-189
10. Tresguerres, M., Buck, J., and Levin, L. (2010) Physiological carbon dioxide, bicarbonate, and pH sensing. *Pflügers Archiv - European Journal Physiology* **460**, 953-964
11. Moroney, J. V., Bartlett, S. G., and Samuelsson, G. (2001) Carbonic anhydrases in plants and algae. *Plant, Cell & Environment* **24**, 141-153
12. Haldane, J. S., and Priestley, J. G. (1905) The regulation of the lung-ventilation. *The Journal of physiology* **32**, 225
13. Bretscher, A. J., Busch, K. E., and de Bono, M. (2008) A carbon dioxide avoidance behavior is integrated with responses to ambient oxygen and food in Caenorhabditis elegans. *PNAS* **105**, 8044-8049
14. Huckstepp, R. T., and Dale, N. (2011) CO_2-dependent opening of an inwardly rectifying K^+ channel. *Pflügers Archiv-European Journal Physiology* **461**, 337-344
15. Richerson, G. (2004) Serotonergic neurons as carbon dioxide sensors that maintain ph homeostasis. *Nature Reviews Neuroscience* **5**, 449-461
16. Demidchik, V., Davenport, R. J., and Tester, M. (2002) Nonselective cation channels in plants. *Annual review of plant biology* **53**, 67-107
17. Sirker, A. A., Rhodes, A., Grounds, R. M., and Bennett, E. D. (2002) Acid−base physiology: the 'traditional' and the 'modern' approaches. *Anaesthesia* **57**, 348-356
18. Pitts, R., and Lotspeich, W. (1946) Bicarbonate and the renal regulation of acid base balanace. *Legacy Content* **147**, 138-154

19. Danielsen, A. A., Parker, M. D., Lee, S., Boron, W. F., Aalkjaer, C., and Boedtkjer, E. (2013) Splice Cassette II of Na+, HCO3− Cotransporter NBCn1 (slc4a7) Interacts with Calcineurin A implications for transporter activity and intracellular pH control during rat artery contractions. *Journal Biological Chemistry* **288**, 8146-8155

20. Luo, M., Sun, L., and Hu, J. (2009) Neural detection of gases-carbon dioxide, oxygen-;in vertebrates and invertebrates. *Current Opinion in Neurobiology* **19**, 354-361

21. Simon, S. A., and Gutknecht, J. (1980) Solubility of carbon dioxide in lipid bilayer membranes and organic solvents. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **596**, 352-358

22. Endeward, V., Musa-Aziz, R., Cooper, G. J., Chen, L. M., Pelletier, M. F., Virkki, L. V., Supuran, C. T., King, L. S., Boron, W. F., and Gros, G. (2006) Evidence that aquaporin 1 is a major pathway for CO_2 transport across the human erythrocyte membrane. *Faseb Journal* **20**, 1974-1981

23. Herrera, M., and Garvin, J. (2011) Aquaporins as gas channels. *Pflügers Archiv - European Journal Physiology* **462**, 623-630

24. Endeward, V., Cartron, J. P., Ripoche, P., and Gros, G. (2008) RhAG protein of the Rhesus complex is a CO_2 channel in the human red cell membrane. *Faseb Journal* **22**, 64-73

25. Arthurs, G., and Sudhakar, M. (2005) Carbon Dioxide Transport. *Continuing Education in Anaesthesia Critical Care* **5**, 207-210

26. Geers, C., and Gros, G. (2000) Carbon Dioxide Transport and Carbonic Anhydrase in Blood and Muscle. *Physiological Reviews* **80**, 681-715

27. Enns, T. (1967) Facilitation by Carbonic Anhydrase of Carbon Dioxide Transport. *Science* **155**, 44-47

28. Slyke, D., and Cullen, G. (1917) The bicarbonate concentration of the blood plasma; its significance, and its determination as a measure of acidosis. *Journal Biological Chemistry* **30**, 289-346

29. Rahman, N., Buck, J., and Levin, L. R. (2013) pH sensing via bicarbonate-regulated “soluble” adenylyl cyclase (sAC). *Frontiers in Physiology* **4**, 343

30. Townsend, P. D., Holliday, P. M., Fenyk, S., Hess, K. C., Gray, M. A., Hodgson, D. R., and Cann, M. J. (2009) Stimulation of mammalian G-protein-responsive adenylyl cyclases by carbon dioxide. *Journal Biological Chemistry* **284**, 784-791

31. Garthwaite, J., and Boulton, C. (1995) Nitric oxide signaling in the central nervous system. *Annual Review Physiology* **57**, 683-706

32. Ando, H., Mizutani, A., Matsu-ura, T., and Mikoshiba, K. (2003) IRBIT, a novel inositol 1, 4, 5-trisphosphate (IP3) receptor-binding protein, is released from the IP3 receptor upon IP3 binding to the receptor. *Journal Biological Chemistry* **278**, 10602-10612

33. Mikoshiba, K. (2007) IP3 receptor/Ca2+ channel: from discovery to new signaling concepts. *Journal of neurochemistry* **102**, 1426-1446

34. Shirakabe, K., Priori, G., Yamada, H., Ando, H., Horita, S., Fujita, T., Fujimoto, I., Mizutani, A., Seki, G., and Mikoshiba, K. (2006) IRBIT, an inositol 1, 4, 5-trisphosphate receptor-binding protein, specifically binds to and activates pancreas-type Na+/HCO3− cotransporter 1 (pNBC1). *PNAS* **103**, 9542-9547

35. Seki, G., Yamada, H., Horita, S., Suzuki, M., Sekine, T., Igarashi, T., and Fujita, T. (2008) Activation and inactivation mechanisms of Na-HCO3 cotransporter NBC1. *Journal of Epithelial Biology & Pharmacology* **1**

36. Nakade, S., Rhee, S., Hamanaka, H., and Mikoshiba, K. (1994) Cyclic AMP-dependent phosphorylation of an immunoaffinity-purified homotetrameric inositol 1, 4, 5-trisphosphate receptor (type I) increases Ca2+ flux in reconstituted lipid vesicles. *Journal Biological Chemistry* **269**, 6735-6742

37. Bruce, J. I. E., Straub, S. V., and Yule, D. I. (2003) Crosstalk between cAMP and Ca2+ signaling in non-excitable cells. *Cell Calcium* **34**, 431-444

38. Sauer, H., Wartenberg, M., and Hescheler, J. (2001) Reactive oxygen species as intracellular messengers during cell growth and differentiation. *Cellular Physiology and Biochemistry* **11**, 173-186

39. Hensley, K., Robinson, K. A., Gabbita, S. P., Salsman, S., and Floyd, R. A. (2000) Reactive oxygen species, cell signaling, and cell injury. *Free Radical Biology and Medicine* **28**, 1456-1462

40. Tsung, A., Klune, J. R., Zhang, X., Jeyabalan, G., Cao, Z., Peng, X., Stolz, D. B., Geller, D. A., Rosengart, M. R., and Billiar, T. R. (2007) HMGB1 release induced by liver ischemia involves Toll-like receptor 4–dependent reactive oxygen species production and calcium-mediated signaling. *The Journal of experimental medicine* **204**, 2913-2923

41. Bolevich, S., Kogan, A. H., Zivkovic, V., Djuric, D., Novikov, A. A., Vorobyev, S. I., and Jakovljevic, V. (2016) Protective role of carbon dioxide (CO2) in generation of reactive oxygen species. *Molecular and cellular biochemistry* **411**, 317-330

42. Taylor, C., and Cummins, E. (2011) Regulation of gene expression by carbon dioxide. *The Journal of Physiology* **589**, 797-803

43. Karakurum, M., Shreeniwas, R., Chen, J., Pinsky, D., Yan, S., Anderson, M., Sunouchi, K., Major, J., Hamilton, T., and Kuwabara, K. (1994) Hypoxic induction of interleukin-8 gene expression in human endothelial cells. *Journal of Clinical Investigation* **93**, 1564

44. Guais, A., Brand, G., Jacquot, L., Karrer, M., Dukan, S., Grévillot, G., Molina, T. J., Bonte, J., Regnier, M., and Schwartz, L. (2011) Toxicity of carbon dioxide: a review. *Chemical research in toxicology* **24**, 2061-2070

45. Cummins, E. P., Oliver, K. M., Lenihan, C. R., Fitzpatrick, S. F., Bruning, U., Scholz, C. C., Slattery, C., Leonard, M. O., McLoughlin, P., and Taylor, C. T. (2010) NF-κB links CO2 sensing to innate immunity and inflammation in mammalian cells. *The Journal of Immunology* **185**, 4439-4445

46. Meier-Schellersheim, M., Xu, X., Angermann, B., Kunkel, E. J., Jin, T., and Germain, R. N. (2006) Key role of local regulation in chemosensing revealed by a new molecular interaction-based modeling method. *PLoS Comput Biol* **2**, e82

47. Putnam, R. W., Filosa, J. A., and Ritucci, N. A. (2004) Cellular mechanisms involved in CO2 and acid signaling in chemosensitive neurons. *American Journal of Physiology-Cell Physiology* **287**, C1493-C1526

48. Huckstepp, R. T. R., Bihi, R. I., Eason, R., Spyer, K. M., Dicke, N., Willecke, K., Marina, N., Gourine, A. V., and Dale, N. (2010) Connexin hemichannel-mediated CO2-dependent release of ATP in the medulla oblongata contributes to central respiratory chemosensitivity. *Journal of Physiology-London* **588**, 3901-3920

49. Buck, J., and Levin, L. R. (2011) Physiological sensing of carbon dioxide/bicarbonate/pH via cyclic nucleotide signaling. *Sensors* **11**, 2112-2128

50. Pattison, R. N., Swamy, J., Mendenhall, B., Hwang, C., and Frohlich, B. T. (2000) Measurement and Control of Dissolved Carbon Dioxide in Mammalian Cell Culture Processes Using an in Situ Fiber Optic Chemical Sensor. *Biotechnology Progress* **16**, 769-774

51. Jones, W., Cayirlioglu, P., Kadow, I., and Vosshall, L. (2007) Two chemosensory receptors together mediate carbon dioxide detection in *Drosophila*. *Nature* **445**, 86-90

52. Sharabi, K., Hurwitz, A., Simon, A. J., Beitel, G. J., Morimoto, R. I., Rechavi, G., Sznajder, J. I., and Gruenbaum, Y. (2009) Elevated CO(2) levels affect development, motility, and fertility and extend life span in Caenorhabditis elegans. *PNAS* **106**, 4024-4029

53. Hallem, E. A., and Sternberg, P. W. (2008) Acute carbon dioxide avoidance in Caenorhabditis elegans. *PNAS* **105**, 8038-8043

54. Guo, D., Zhang, J. J., and Huang, X.-Y. (2009) Stimulation of Guanylyl Cyclase-D by Bicarbonate. *Biochemistry* **48**, 4417-4422

55. Sun, L., Wang, H., Hu, J., Han, J., Matsunami, H., and Luo, M. (2009) Guanylyl cyclase-D in the olfactory CO_2 neurons is activated by bicarbonate. *PNAS* **106**, 2041-2046

56. Evans, M., Buchanan, B. B., and Arnon, D. I. (1966) A new ferredoxin-dependent carbon reduction cycle in a photosynthetic bacterium. *PNAS* **55**, 928-934

57. Raines, C. A. (2003) The Calvin cycle revisited. *Photosynthesis Res* **75**, 1-10

58. Field, C. B. (2012) *Managing the risks of extreme events and disasters to advance climate change adaptation: special report of the intergovernmental panel on climate change*, Cambridge University Press

59. Kimball, B. A. (1983) Carbon dioxide and agricultural yield: an assemblage and analysis of 430 prior observations. *Agronomy journal* **75**, 779-788

60. Kimball, B. A. (2016) Crop responses to elevated CO_2 and interactions with H_2O, N, and temperature. *Current Opinion in Plant Biology* **31**, 36-43

61. Long, S. P., Ainsworth, E. A., Leakey, A. D., Nösberger, J., and Ort, D. R. (2006) Food for thought: lower-than-expected crop yield stimulation with rising CO_2 concentrations. *Science* **312**, 1918-1921

62. Haworth, M., Hoshika, Y., and Killi, D. (2016) Has the Impact of Rising CO_2 on Plants been Exaggerated by Meta-Analysis of Free Air CO_2 Enrichment Studies? *Frontier Plant Science* **7**, 1153

63. Long, S. P., Ainsworth, E. A., Rogers, A., and Ort, D. R. (2004) Rising atmosphereic carbon dioxide: Plants FACE the Future*. *Annual Review of Plant Biology* **55**, 591-628

64. Drake, B. G., Gonzàlez-Meler, M. A., and Long, S. P. (1997) More efficient plants: a consequence of rising atmospheric CO_2? *Annual review of plant biology* **48**, 609-639

65. Stiling, P., and Cornelissen, T. (2007) How does elevated carbon dioxide (CO_2) affect plant–herbivore interactions? A field experiment and meta-

analysis of CO_2-mediated changes on plant chemistry and herbivore performance. *Global Change Biology* **13**, 1823-1842

66. Watson-Lazowski, A., Lin, Y., Miglietta, F., Edwards, R. J., Chapman, M. A., and Taylor, G. (2016) Plant adaptation or acclimation to rising CO_2 ? Insight from first multigenerational RNA-Seq transcriptome. *Global Chang Biology*

67. Song, C.-P., Guo, Y., Qiu, Q., Lambert, G., Galbraith, D. W., Jagendorf, A., and Zhu, J.-K. (2004) A probable Na^+ (K^+)/H^+ exchanger on the chloroplast envelope functions in pH homeostasis and chloroplast development in Arabidopsis thaliana. *PNAS***101**, 10211-10216

68. Stitt, M., and Krapp, A. (1999) The interaction between elevated carbon dioxide and nitrogen nutrition: the physiological and molecular background. *Plant, Cell & Environment* **22**, 583-621

69. Uehlein, N., Lovisolo, C., Siefritz, F., and Kaldenhoff, R. (2003) The tobacco aquaporin NtAQP1 is a membrane CO_2 pore with physiological functions. *Nature* **425**, 734-737

70. Kim, T.-H., Böhmer, M., Hu, H., Nishimura, N., and Schroeder, J. I. (2010) Guard cell signal transduction network: advances in understanding abscisic acid, CO_2, and Ca^{2+} signaling. *Annual review of plant biology* **61**, 561

71. Xue, S., Hu, H., Ries, A., Merilo, E., Kollist, H., and Schroeder, J. I. (2011) Central functions of bicarbonate in S-type anion channel activation and OST1 protein kinase in CO_2 signal transduction in guard cell. *The EMBO journal* **30**, 1645-1658

72. Vahisalu, T., Kollist, H., Wang, Y.-F., Nishimura, N., Chan, W.-Y., Valerio, G., Lamminmaki, A., Brosche, M., Moldau, H., Desikan, R., Schroeder, J. I., and Kangasjarvi, J. (2008) SLAC1 is required for plant guard cell S-type anion channel function in stomatal signalling. *Nature* **452**, 487-491

73. Schroeder, J. I., Allen, G. J., Hugouvieux, V., Kwak, J. M., and Waner, D. (2001) Guard cell signal transduction. *Annual review of plant biology* **52**, 627-658

74. Cotelle, V., and Leonhardt, N. (2015) 14-3-3 Proteins in Guard Cell Signaling. *Frontier Plant Science* **6**, 1210

75. Azzam, Z., Sharabi, K., Guetta, J., Bank, E., and Gruenbaum, Y. (2010) The physiological and molecular effect of elevated CO_2 levels. *Cell Cycle* **9**, 1528-1532

76. Tang, X. D., Santarelli, L. C., Heinemann, S. H., and Hoshi, T. (2004) Metabolic regulation of potassium channels. *Annual Review Physiology* **66**, 131-159

77. Brearley, J., Venis, M. A., and Blatt, M. R. (1997) The effect of elevated CO_2 concentrations on K^+ and anion channels of Vicia faba L. guard cells. *Planta* **203**, 145-154

78. Xu, Z., Jiang, Y., Jia, B., and Zhou, G. (2016) Elevated-CO_2 Response of Stomata and Its Dependence on Environmental Factors. *Frontier Plant Science* **7**, 657

79. Negi, J., Matsuda, O., Nagasawa, T., Oba, Y., Takahashi, H., Kawai-Yamada, M., Uchimiya, H., Hashimoto, M., and Iba, K. (2008) CO_2 regulator SLAC1 and its homologues are essential for anion homeostasis in plant cells. *Nature* **452**, 483-486

80. Kollipara, L., and Zahedi, R. P. (2013) Protein carbamylation: in vivo modification or in vitro artefact? *Proteomics* **13(6)**, 941-944

81. Hampe, E. M., and Rudkevich, D. M. (2003) Exploring reversible reactions between CO_2 and amines. *Tetrahedron* **59**, 9619-9625

82. Jörgensen, I., and Stiles, W. (1917) Carbon Assimilation. *New Phytologist* **16**, 77-104

83. Lorimer, G., and Miziorko, H. (1980) Carbamate formation on the e-amino group of a lysyl residue as the basis for the activation of Ribulosebisphosphate Carboxylase by CO_2 and Mg^{2+}. *Biochemistry* **19**, 5321-5328

84. Terrier, P., and Douglas, D. J. (2010) Carbamino Group Formation with Peptides and Proteins Studied by Mass Spectrometry. *Journal of the American Society for Mass Spectrometry* **21**, 1500-1505

85. Jimenez-Morales, D., Adamian, L., Shi, D., and Liang, J. (2014) Lysine carboxylation: unveiling a spontaneous post-translational modification. *Acta Crystallographica Section D: Biological Crystallography* **70**, 48-57

86. Morrow, J., Keim, P., and Gurd, F. (1974) CO_2 adducts of certain amino acids, peptides, and sperm whale myoglobin studied by carbon 13 and proton nucleear magnetic resonance. *Journal Biological Chemistry* **249**, 7484-7494

87. Stadie, W. C., and O'Brien, H. (1936) The carbamate equilibrium: I. the equilibrium of amino acids, carbon dioxide and carbamates in aqueous solution; with a note on the ferguson-roughton carbamate method. *Journal Biological Chemistry* **112**, 723-758

88. Abraham, S., Kobayashi, T., John Solaro, R., and Gaponenko, V. (2009) Differences in lysine pKa values may be used to improve NMR signal dispersion in reductively methylated proteins. *Journal Biomolecular NMR* **43**, 239-246

89. Hackling, M. W., and Garnett, P. J. (1985) Misconceptions of chemical equilibrium. *European Journal of Science Education* **7**, 205-214

90. Gros, G., and Bauer, C. (1978) High pK value of the N-terminal amino group of the γ-chain causes low CO_2 binding of human fetal hemoglobin. *Biochemical and Biophysical Research Communications* **80**, 56-62

91. Hsia, C. C. W. (1998) Respiratory Function of Hemoglobin. *New England Journal of Medicine* **338**, 239-248

92. Matthew, J., Morrow, J., Wittebort, R., and Gurd, F. (1977) Quantittive determination of caramino adductsd of alpha and beta chains in human adult hemoglobin in presence and absence of carbon monoxide and 2,3-diphosphoglycerate. *Journal Biological Chemistry* **252**, 2234-2244

93. Vandegriff, K., Benazzi, L., Ripamonti, M., Perrella, M., Tellier, Y., Zegna, A., and Winslow, R. (1991) Carbon dioxide binding to human hemoglobin cross-linked between the alpha chains. *Journal Biological Chemistry* **266**, 2697-2700

94. Vallet, B., Teboul, J.-L., Cain, S., and Curtis, S. (2000) Venoarterial CO_2 difference during regional ischemic or hypoxic hypoxia. *Journal of Applied Physiology* **89**, 1317-1321

95. Lorimer, G. (1983) Carbon dioxide and carbamate formation: the makings of a biochemical control system. *Trends in Biochemical Sciences* **8**, 65-68

96. Meigh, L., Cook, D., Zhang, J., and Dale, N. (2015) Rational design of new NO and redox sensitivity into connexin26 hemichannels. *Open Biology* **5**, 9

97. Meigh, L., Greenhalgh, S. A., Rodgers, T. L., Cann, M. J., Roper, D. I., and Dale, N. (2013) CO_2 directly modulates connexin 26 by formation of carbamate bridges between subunits. *Elife* **2**

98. Huckstepp, R. T. R., and Dale, N. (2011) Redefining the components of central CO_2 chemosensitivity – towards a better understanding of mechanism. *The Journal of Physiology* **589**, 5561-5579

99. de Wolf, E., Cook, J., and Dale, N. (2017) Evolutionary adaptation of the sensitivity of connexin26 hemichannels to CO2. in *Proc. R. Soc. B*, The Royal Society **284,** 20162723

100. Stec, B. (2012) Structural mechanism of RuBisCO activation by carbamylation of the active site lysine. *PNAS* **109**, 18785-18790

101. Drummond, M., Wilson, A., and Cundari, T. (2012) Nature of protein-CO_2 interactions as elucidated via molecular dynamics. *The Journal of Physical Chemistry* **116**, 11578-11593

102. Andersson, I. (2008) Catalysis and regulation in Rubisco. *Journal Experimental Botany* **59**, 1555-1568

103. Drummond, M. L., Wilson, A. K., and Cundari, T. R. (2012) Carbon Dioxide Migration Pathways in Proteins. *The Journal of Physical Chemistry Letters* **3**, 830-833

104. Hartman, F. C., and Harpel, M. (1994) Structure, Function, Regulation and Assembly of D-Ribulose-1,5-Bisphosphate Carboxylase/Oxygenase. *Annual Review of Biochemistry* **63**, 197-232

105. Lorimer, G. (1979) Evidence for the existence of discrete activator and substrate sites for CO_2 on Ribulose-1,5-bisphosphate carboxylase. *Journal Biological Chemistry* **254**, 5599-5601

106. Knight, S., Andersson, I., and Branden, C.-I. (1990) Crystallographic analysis of Ribulose 1,5-bisphosphate carboxylase from spinach at 2.4 A resolution. *Journal of Molecular Biology* **215**, 113-160

107. Smith, H. B., Larimer, F. W., and Hartman, F. C. (1988) Subtle alteration of the active site of ribulose bisphosphate carboxylase/oxygenase by concerted site-directed mutagenesis and chemical modification. *Biochemical and Biophysical Research Communications* **152**, 579-584

108. Yamaguchi, K., and Hausinger, R. P. (1997) Substitution of the Urease Active Site Carbamate by Dithiocarbamate and Vanadate. *Biochemistry* **36**, 15118-15122

109. Pearson, M. A., Schaller, R. A., Michel, L. O., Karplus, P. A., and Hausinger, R. P. (1998) Chemical rescue of Klebsiella aerogenes urease variants lacking the carbamylated-lysine nickel ligand. *Biochemistry* **37**, 6214-6220

110. Morollo, A., Petsko, G., and Ringe, D. (1999) Structure of a Michaelis complex analogue: Propionate binds in the substrate carboxylate site of Alanine Racemase. *Biochemistry* **38**, 3293-3301

111. Samols, D., Thornton, C. G., Murtif, V. L., Kumar, G. K., Haase, F. C., and Wood, H. G. (1988) Evolutionary conservation among biotin enzymes. *Journal Biological Chemistry* **263**, 6461-6464

112. Hall, P., Zheng, R., Antony, L., Pusztai-Carey, M., Carey, P., and Yee, V. (2004) Transcarboxylase 5S structures: assembly and catalytic mechanism of a multienzyme complex subunit. *The EMBO Journal* **23**, 3621-3631

113. Wood, H. G., and Barden, R. E. (1977) Biotin enzymes. *Annual Review of Biochemistry* **46**, 385-413

114. Maveyraud, L., Golemi, D., Kotra, L. P., Tranier, S., Vakulenko, S., Mobashery, S., and Samama, J.-P. (2000) Insights into Class D β-Lactamases Are Revealed by the Crystal Structure of the OXA10 Enzyme from Pseudomonas aeruginosa. *Structure* **8**, 1289-1298

115. Golemi, D., Maveyraud, L., Vakulenoko, S., Samama, J.-P., and Mobashery, S. (2001) Critical involvement of a carbamylated lysine in catalytic function of class D beta-lactamases *PNAS* **98**, 14280-14285

116. Bigley, A. N., and Raushel, F. M. (2013) Catalytic Mechanisms for Phosphotriesterases. *Biochimica et biophysica acta* **1834**, 443-453

117. Benning, M. M., Kuo, J. M., Raushel, F. M., and Holden, H. M. (1995) Three-dimensional structure of the binuclear metal center of phosphotriesterase. *Biochemistry* **34**, 7973-7978

118. Jackson, C., Kim, H.-K., Carr, P. D., Liu, J.-W., and Ollis, D. L. (2005) The structure of an enzyme–product complex reveals the critical role of a terminal hydroxide nucleophile in the bacterial phosphotriesterase mechanism. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* **1752**, 56-64

119. Wittebort, R. J., Hayes, D. F., Rothgeb, T. M., and Gurd, R. (1978) The quantitation of carbamino adduct formation of angiotensin II and bradykinin. *Biophysical journal* **24**, 765

120. Worley, J. F., 3rd, French, R. J., and Krueger, B. K. (1986) Trimethyloxonium modification of single batrachotoxin-activated sodium channels in planar bilayers. Changes in unit conductance and in block by saxitoxin and calcium. *The Journal of general physiology* **87**, 327-349

121. Aoyama, T., and Shioiri, T. (1990) Trimethylsilyldiazomethane : a convenient reagent for the o-methylation of alcohols. *Tetrahedron Letters* **31**, 5507-5508

122. Ito, Y., and Ushitora, H. (2006) Trapping of carbamic acid species with (trimethylsilyl)diazomethane. *Tetrahedron* **62**, 226-235

123. King, G. J., Gazzola, C., Blakeley, R. L., and Zerner, B. (1986) Triethyloxonium tetrafluoroborate as an ethylating agent in aqueous solution. *Inorganic Chemistry* **25**, 1078-1078

124. McGarrity, J. F., and Smyth, T. (1980) Hydrolysis of diazomethane-kinetics and mechanism. *Journal of the American Chemical Society* **102**, 7303-7308

125. Reed, J. K., and Raftery, M. A. (1976) Properties of tetrodotoxin binding component in plasma-membranes isolated from electrophorus-electricus. *Biochemistry* **15**, 944-953

126. Herskovits, T. T., Gadegbeku, B., and Jaillet, H. (1970) On the structural stability and solvent denaturation of proteins I. Denaturation by the alcohols and glycols. *Journal Biolgical Chemistry* **245**, 2588-2598

127. Drummond, M., Wilson, A., and Cundari, T. (2010) Toward greener carbon capture technologies: A pharmacophore-based approach to predict CO_2 binding sites in proteins. *Energy Fuels* **24**, 1464-1470

128. Pörtner, H., Langenbuch, M., and Reipschläger, A. (2004) Biological Impact of Elevated Ocean CO_2 Concentrations: Lessons from Animal Physiology and Earth History. *Journal of Oceanography* **60**, 705-718

129. Riebesell, U., Zondervan, I., Rost, B., Tortell, P., Zeebe, R., and Morel, F. (2000) Reduced calcification of marine plankton in response to increased atmospheric CO_2. *Nature* **407**, 364-367

130. O'Neill, L. A. J., and Kaltschmidt, C. (1997) NF-kB: a crucial transcription factor for glial and neuronal cell function. *Trends in Neurosciences* **20**, 252-258

131. Leakey, A. D., Bishop, K. A., and Ainsworth, E. A. (2012) A multi-biome gap in understanding of crop and ecosystem responses to elevated CO_2. *Current opinion in plant biology* **15**, 228-236

132. Leakey, A. D., Ainsworth, E. A., Bernacchi, C. J., Rogers, A., Long, S. P., and Ort, D. R. (2009) Elevated CO_2 effects on plant carbon, nitrogen, and water relations: six important lessons from FACE. *Journal Experimental Botany* **60**, 2859-2876

133. Myers, S. S., Zanobetti, A., Kloog, I., Huybers, P., Leakey, A. D., Bloom, A. J., Carlisle, E., Dietterich, L. H., Fitzgerald, G., Hasegawa, T., Holbrook, N. M., Nelson, R. L., Ottman, M. J., Raboy, V., Sakai, H., Sartor, K. A., Schwartz, J., Seneweera, S., Tausz, M., and Usui, Y. (2014) Increasing CO_2 threatens human nutrition. *Nature* **510**, 139-142

134. Kant, S., Seneweera, S., Rodin, J., Materne, M., Burch, D., Rothstein, S. J., and Spangenberg, G. (2012) Improving yield potential in crops under elevated CO_2: Integrating the photosynthetic and nitrogen utilization efficiencies. *Frontiers in Plant Science* **3**, 162

135. Rappsilber, J., Mann, M., and Ishihama, Y. (2007) Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nature Protocols* **2**, 1896-1906

136. Craig, R., Cortens, J. P., and Beavis, R. C. (2004) Open Source System for Analyzing, Validating, and Storing Protein Identification Data. *Journal of Proteome Research* **3**, 1234-1242

137. Cox, J., Neuhauser, N., Michalski, A., Scheltema, R. A., Olsen, J. V., and Mann, M. (2011) Andromeda: A Peptide Search Engine Integrated into the MaxQuant Environment. *Journal of Proteome Research* **10**, 1794-1805

138. Smith, R. J., and Bienz, S. (2004) Towards Functionalized Silicon-Containing α-Amino Acids: Asymmetric Syntheses of Sila Analogs of Homoserine and Homomethionine. *Helvetica Chimica Acta* **87**, 1681-1696

139. Montalbetti, C. A. G. N., and Falque, V. (2005) Amide bond formation and peptide coupling. *Tetrahedron* **61**, 10827-10852

140. Nulwala, H. B., Tang, C. N., Kail, B. W., Damodaran, K., Kaur, P., Wickramanayake, S., Shi, W., and Luebke, D. R. (2011) Probing the structure-property relationship of regioisomeric ionic liquids with click chemistry. *Green Chemistry* **13**, 3345-3349

141. Presser, A., and Hüfner, A. (2004) Trimethylsilyldiazomethane – A Mild and Efficient Reagent for the Methylation of Carboxylic Acids and Alcohols in Natural Products. *Monatshefte für Chemie* **135**, 1015-1022

142. Myers, E. L., and Raines, R. T. (2009) A Phosphine-Mediated Conversion of Azides into Diazo Compounds. *Angewandte Chemie-International Edition* **48**, 2359-2363

143. Morrow, J. S., Keim, P., Visscher, R. B., Marshall, R. C., and Gurd, F. R. N. (1973) Interaction of $^{13}CO_2$ and Bicarbonate with Human Hemoglobin Preparations. *PNAS* **70**, 1414-1418

144. O'Leary, M. H., Jaworski, R. J., and Hartman, F. C. (1979) ^{13}C nuclear magnetic resonance study of the CO_2 activation of ribulosebisphosphate carboxylase from Rhodospirillum rubrum. *PNAS* **76**, 673-675

145. Kilmartin, J. V. (1976) Interaction of Haemoglobin with protons, CO2 and 2,3-diphosphoglycerate. *British Medical Bulletin* **32**, 209-212

146. Raber, D. J., Gariano, P., Brod, A. O., Gariano, A., Guida, W. C., Guida, A. R., and Herbst, M. D. (1979) Esterification of carboxylic acids with trialkyloxonium salts. *The Journal of Organic Chemistry* **44**, 1149-1154

147. Kilmartin, J. V., and Rossi-Bernardi, L. (1971) The binding of carbon dioxide by horse haemoglobin. *Biochemical Journal* **124**, 31-45

148. Schwartz, D. P., and Bright, R. S. (1974) A column procedure for the esterification of organic acids with diazomethane at the microgram level. *Analytical Biochemistry* **61**, 271-274

149. Duthaler, R. O., Foerster, H. G., and Roberts, J. D. (1978) Nitrogen-15 and carbon-13 nuclear magnetic resonance spectra of diazo and diazonium compounds. *Journal of the American Chemical Society* **100**, 4974-4979

150. Kühnel, E., Laffan, D. D. P., Lloyd-Jones, G. C., Martínez del Campo, T., Shepperson, I. R., and Slaughter, J. L. (2007) Mechanism of Methyl Esterification of Carboxylic Acids by Trimethylsilyldiazomethane. *Angewandte Chemie International Edition* **46**, 7075-7078

151. Maurya, R. A., Park, C. P., Lee, J. H., and Kim, D.-P. (2011) Continuous In Situ Generation, Separation, and Reaction of Diazomethane in a Dual-Channel Microreactor. *Angewandte Chemie International Edition* **50**, 5952-5955

152. Peterson, S. L., Stucka, S. M., and Dinsmore, C. J. (2010) Parallel Synthesis of Ureas and Carbamates from Amines and CO_2 under Mild Conditions. *Organic Letters* **12**, 1340-1343

153. Shioiri, T., Aoyama, T., and Mori, S. (1990) Trimethylsilyldiazomethane - (silane, (diazomethyl)trimethyl). *Organic Syntheses* **68**, 1-7

154. Glastrup, J. (1998) Diazomethane preparation for gas chromatographic analysis. *Journal of Chromatography A* **827**, 133-136

155. Aldai, N., Murray, B. E., Nájera, A. I., Troy, D. J., and Osoro, K. (2005) Derivatization of fatty acids and its application for conjugated linoleic acid studies in ruminant meat lipids. *Journal of the Science of Food and Agriculture* **85**, 1073-1083

156. Aoyama, T., and Shioiri, T. (1980) New methods and reagents in organic synthesis. 8. Trimethylsilyldiazomethane. A new, stable, and safe reagent for the classical arndt-eistert synthesis. *Tetrahedron Letters* **21**, 4461-4462

157. Park, Y., Albright, K. J., Cai, Z. Y., and Pariza, M. W. (2001) Comparison of Methylation Procedures for Conjugated Linoleic Acid and Artifact Formation by Commercial (Trimethylsilyl)diazomethane. *Journal of Agricultural and Food Chemistry* **49**, 1158-1164

158. Mattos, C., and Ringe, D. (2001) Proteins in organic solvents. *Current opinion in structural biology* **11**, 761-764

159. Stella, V. J., and Nti-Addae, K. W. (2007) Prodrug strategies to overcome poor water solubility. *Advanced Drug Delivery Reviews* **59**, 677-694

160. Mori, S., Sakai, I., Aoyama, T., and Shioiri, T. (1982) New methods and reagents in organic-synthesis .28. a convenient and efficient preparation of trimethylsilyldiazomethane (TMSCHN2) using diphenyl phosphorazidate (DPPA). *Chemical & Pharmaceutical Bulletin* **30**, 3380-3382

161. Klapars, A., and Buchwald, S. L. (2002) Copper-catalyzed halogen exchange in aryl halides: An aromatic Finkelstein reaction. *Journal of the American Chemical Society* **124**, 14844-14845

162. Kubota, K., Yamamoto, E., and Ito, H. (2013) Copper(I)-Catalyzed Borylative exo-Cyclization of Alkenyl Halides Containing Unactivated Double Bond. *Journal of the American Chemical Society* **135**, 2635-2640

163. Dongol, K. G., Koh, H., Sau, M., and Chai, C. L. L. (2007) Iron-Catalysed sp3–sp3 Cross-Coupling Reactions of Unactivated Alkyl Halides with Alkyl Grignard Reagents. *Advanced Synthesis & Catalysis* **349**, 1015-1018

164. Trent, J. S. (1984) Ruthenium tetraoxide staining of polymers: new preparative methods for electron microscopy. *Macromolecules* **17**, 2930-2931

165. Carlsen, P. H. J., Katsuki, T., Martin, V. S., and Sharpless, K. B. (1981) A greatly improved procedure for ruthenium tetroxide catalyzed oxidations of organic compounds. *The Journal of Organic Chemistry* **46**, 3936-3938

166. Fox M. A., J. K. W. (2003) *Organic Chemistry third edition*, Jones and Bartlett Publishers, Sudbury, Massachusetts

167. Pfitzner, K. E., and Moffatt, J. G. (1965) Sulfoxide-Carbodiimide Reactions. II. Scope of the Oxidation Reaction1. *Journal of the American Chemical Society* **87**, 5670-5678

168. Bender, M. L. (1960) Mechanisms of Catalysis of Nucleophilic Reactions of Carboxylic Acid Derivatives. *Chemical Reviews* **60**, 53-113

169. Tilstam, U., and Weinmann, H. (2002) Activation of Mg Metal for Safe Formation of Grignard Reagents on Plant Scale. *Organic Process Research & Development* **6**, 906-910

170. Hughes, E. D., and Ingold, C. K. (1935) 55. Mechanism of substitution at a saturated carbon atom. Part IV. A discussion of constitutional and solvent effects on the mechanism, kinetics, velocity, and orientation of substitution. *Journal of the Chemical Society (Resumed)*, 244-255

171. Nishiyama, K., and Tanaka, N. (1983) Synthesis and reactions of trimethylsilylmethyl azide. *Journal of the Chemical Society-Chemical Communications*, 1322-1323

172. Taber, D. F., Ruckle Jr, R. E., and Hennessy, M. J. (1986) Mesyl azide: A superior reagent for diazo transfer. *The Journal of Organic Chemistry* **51**, 4077-4078

173. Titz, A., Radic, Z., Schwardt, O., and Ernst, B. (2006) A safe and convenient method for the preparation of triflyl azide, and its use in diazo transfer reactions to primary amines. *Tetrahedron letters* **47**, 2383-2385

174. Scriven, E. F. V., and Turnbull, K. (1988) Azides: their preparation and synthetic uses. *Chemical Reviews* **88**, 297-368

175. Collins, C. J. (1971) Reactions of primary aliphatic amines with nitrous acid. *Accounts of Chemical Research* **4**, 315-322

176. Ye, T., and McKervey, M. A. (1994) Organic Synthesis with .alpha.-Diazo Carbonyl Compounds. *Chemical Reviews* **94**, 1091-1160

177. Wilds, A. L., and Meader, A. L. (1948) The use of higher diazohydrocarbons in the arndt-eistert synthesis. *The Journal of Organic Chemistry* **13**, 763-779

178. Marshall, J. L., Erickson, K. C., and Folsom, T. K. (1970) The esterification of carboxylic acids using a boron trifluoride-etherate-alcohol reagent. *Tetrahedron Letters* **11**, 4011-4012

179. Lillington, J. M., Trafford, D. J. H., and Makin, H. L. J. (1981) A rapid and simple method for the esterification of fatty acids and steroid carboxylic acids prior to gas-liquid chromatography. *Clinica Chimica Acta* **111**, 91-98

180. Lamoureux, G., and Agüero, C. (2009) A comparison of several modern alkylating agents. *Arkivoc* **1**, 251-264

181. Liebich, H. M., and Gesele, E. (1999) Profiling of organic acids by capillary gas chromatography–mass spectrometry after direct methylation in urine using trimethyloxonium tetrafluoroborate. *Journal of Chromatography A* **843**, 237-245

182. Pappone, P. A., and Barchfeld, G. L. (1990) Modifications of single acetylcholine-activated channels in BC3H-1 cells. Effects of trimethyloxonium and pH. *The Journal of general physiology* **96**, 1-22

183. MacKinnon, R., and Miller, C. (1989) Functional modification of calcium-activated potassium channel by trimethyloxonium. *Biochemistry* **28**, 8087-8092

184. Paterson, A. K., and Knowles, J. R. (1972) The Number of Catalytically Essential Carboxyl Groups in Pepsin. *European Journal of Biochemistry* **31**, 510-517

185. Parsons, S. M., Jao, L., Dahlquist, F. W., Borders, C. L., Groff, T., Racs, J., and Raftery, M. A. (1969) Nature of amino acid side chains which are critical for the activity of lysozyme. *Biochemistry* **8**, 700-712

186. Kiessling, A. A., S. (2003) *Human Embryonic Stem Cells: An Introduction to the Science and Therapeutic Potential*, Jones and Bartlett Publishers, Canada

187. Zhang, M., and Vogel, H. J. (1993) Determination of the side chain pKa values of the lysine residues in calmodulin. *Journal Biology Chemistry* **268**, 22420-22428

188. Dewick, P. (2013) *Essentials of Organic Chemistry*, John Wiley & Sons, West Sussex, England

189. Madshus, I. H. (1988) Regulation of intracellular pH in eukaryotic cells. *Biochemical Journal* **250**, 1-8

190. Deutsch, C., Taylor, J. S., and Wilson, D. F. (1982) Regulation of Intracellular pH by Human Peripheral Blood Lymphocytes as Measured by 19F NMR. *PNAS* **79**, 7944-7948

191. Morrow, J., Matthew, J., Wittebort, R., and Gurd, F. (1976) Carbon 13 resonance of $^{13}CO_2$ carbamino adducts of alpha and beta chains in human adult hemoglobin. *Journal Biolgical Chemistry* **251**, 477-484

192. Annesley, T. M. (2003) Ion Suppression in Mass Spectrometry. *Clinical Chemistry* **49**, 1041-1044

193. Issaq, H. J., Chan, K. C., Janini, G. M., Conrads, T. P., and Veenstra, T. D. (2005) Multidimensional separation of peptides for effective proteomic analysis. *Journal of Chromatography B* **817**, 35-47

194. Nørregaard Jensen, O. (2004) Modification-specific proteomics: characterization of post-translational modifications by mass spectrometry. *Current Opinion in Chemical Biology* **8**, 33-41

195. Roepstorff, P. (1997) Mass spectrometry in protein studies from genome to function. *Current Opinion in Biotechnology* **8**, 6-13

196. Shevchenko, A., Wilm, M., Vorm, O., and Mann, M. (1996) Mass spectrometric sequencing of proteins from silver-stained polyacrylamide gels. *Analytical chemistry* **68**, 850-858

197. Zhu, X., and Papayannopoulos, I. A. (2003) Improvement in the Detection of Low Concentration Protein Digests on a MALDI TOF/TOF Workstation by Reducing α-Cyano-4-hydroxycinnamic Acid Adduct Ions. *Journal of Biomolecular Techniques* **14**, 298-307

198. Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F., and Whitehouse, C. M. (1989) Electrospray ionization for mass spectrometry of large biomolecules. *Science* **246**, 64-71

199. Olsen, J. V., Ong, S.-E., and Mann, M. (2004) Trypsin Cleaves Exclusively C-terminal to Arginine and Lysine Residues. *Molecular & Cellular Proteomics* **3**, 608-614

200. Perrella, M., and Rossi-Bernardi, L. (1980) The Determination of CO_2 Bound to Hemoglobin as Carbamate. in *Biophysics and Physiology of Carbon Dioxide* (Bauer, C., Gros, G., and Bartels, H. eds.), Springer Berlin Heidelberg. pp 75-83

201. Hayashi, M., and Nishimura, M. (2006) Arabidopsis thaliana—A model organism to study plant peroxisomes. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **1763**, 1382-1391

202. Evans, J. R., and Von Caemmerer, S. (1996) Carbon dioxide diffusion inside leaves. *Plant Physiology* **110**, 339

203. Bensimon, A., Heck, A. J. R., and Aebersold, R. (2012) Mass Spectrometry–Based Proteomics and Network Biology. *Annual Review of Biochemistry* **81**, 379-405

204. Zhang, Y., Fonslow, B. R., Shan, B., Baek, M.-C., and Yates, J. R. (2013) Protein Analysis by Shotgun/Bottom-up Proteomics. *Chemical reviews* **113**, 2343-2394

205. Xu, C., and Ma, B. (2006) Software for computational peptide identification from MS–MS data. *Drug Discovery Today* **11**, 595-600

206. Rappsilber, J., and Mann, M. (2002) What does it mean to identify a protein in proteomics? *Trends in Biochemical Sciences* **27**, 74-78

207. Gupta, N., Hixson, K. K., Culley, D. E., Smith, R. D., and Pevzner, P. A. (2010) Analyzing protease specificity and detecting in vivo proteolytic events using tandem mass spectrometry. *Proteomics* **10**, 2833-2844

208. Steen, H., and Mann, M. (2004) The abc's (and xyz's) of peptide sequencing. *Nature Review Molecular Cell Biology* **5**, 699-711

209. Kopec, K. K., Bozyczko-Coyne, D., and Williams, M. (2005) Target identification and validation in drug discovery: the role of proteomics. *Biochemical Pharmacology* **69**, 1133-1139

210. Heazlewood, J. L. (2011) The green proteome: challenges in plant proteomics. *Frontiers in plant science* **2**, 6

211. Thakur, S. S., Geiger, T., Chatterjee, B., Bandilla, P., Fröhlich, F., Cox, J., and Mann, M. (2011) Deep and Highly Sensitive Proteome Coverage by LC-MS/MS Without Prefractionation. *Molecular & Cellular Proteomics* **10**

212. Hebert, A. S., Richards, A. L., Bailey, D. J., Ulbrich, A., Coughlin, E. E., Westphall, M. S., and Coon, J. J. (2014) The One Hour Yeast Proteome. *Molecular & Cellular Proteomics* **13**, 339-347

213. Schwacke, R., Schneider, A., van der Graaff, E., Fischer, K., Catoni, E., Desimone, M., Frommer, W. B., Flügge, U.-I., and Kunze, R. (2003) ARAMEMNON, a novel database for Arabidopsis integral membrane proteins. *Plant Physiology* **131**, 16-26

214. Kleffmann, T., Russenberger, D., von Zychlinski, A., Christopher, W., Sjölander, K., Gruissem, W., and Baginsky, S. (2004) The Arabidopsis thaliana Chloroplast Proteome Reveals Pathway Abundance and Novel Protein Functions. *Current Biology* **14**, 354-362

215. Millar, A. H., Sweetlove, L. J., Giegé, P., and Leaver, C. J. (2001) Analysis of the Arabidopsis mitochondrial proteome. *Plant Physiology* **127**, 1711-1727

216. Mann, M., and Jensen, O. N. (2003) Proteomic analysis of post-translational modifications. *Nature biotechnology* **21**, 255-261

217. Wu, X., Oh, M.-H., Schwarz, E. M., Larue, C. T., Sivaguru, M., Imai, B. S., Yau, P. M., Ort, D. R., and Huber, S. C. (2011) Lysine acetylation is a widespread protein modification for diverse proteins in Arabidopsis. *Plant physiology* **155**, 1769-1778

218. Nakagami, H., Sugiyama, N., Mochida, K., Daudi, A., Yoshida, Y., Toyoda, T., Tomita, M., Ishihama, Y., and Shirasu, K. (2010) Large-scale comparative phosphoproteomics identifies conserved phosphorylation sites in plants. *Plant physiology* **153**, 1161-1174

219. Viner, R., Zhang, T., Peterman, S., Zabrouskov, V., and Jose, S. (2007) Advantages of the LTQ Orbitrap for protein identification in complex digests. *Thermo Scientific Application Note AN* **386**, 12

220. Jones, C. G., Daniel Hare, J., and Compton, S. J. (1989) Measuring plant protein with the Bradford assay. *Journal of Chemical Ecology* **15**, 979-992

221. Bradford, M. M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical biochemistry* **72**, 248-254

222. Compton, S. J., and Jones, C. G. (1985) Mechanism of dye response and interference in the Bradford protein assay. *Analytical biochemistry* **151**, 369-374

223. Smith, P. K., Krohn, R. I., Hermanson, G. T., Mallia, A. K., Gartner, F. H., Provenzano, M. D., Fujimoto, E. K., Goeke, N. M., Olson, B. J., and Klenk, D. C. (1985) Measurement of protein using bicinchoninic acid. *Analytical Biochemistry* **150**, 76-85

224. Gundry, R. L., White, M. Y., Murray, C. I., Kane, L. A., Fu, Q., Stanley, B. A., and Van Eyk, J. E. (2009) Preparation of Proteins and Peptides for Mass Spectrometry Analysis in a Bottom-Up Proteomics Workflow. *Current protocols in molecular biology / edited by Frederick M. Ausubel,* Unit10.25-Unit10.25

225. Widjaja, I., Naumann, K., Roth, U., Wolf, N., Mackey, D., Dangl, J. L., Scheel, D., and Lee, J. (2009) Combining subproteome enrichment and Rubisco depletion enables identification of low abundance proteins differentially regulated during plant defense. *Proteomics* **9**, 138-147

226. Feller, U., Anders, I., and Mae, T. (2008) Rubiscolytics: fate of Rubisco after its enzymatic function in a cell is terminated. *Journal Experimental Botany* **59**, 1615-1624

227. Jiang, L., He, L., and Fountoulakis, M. (2004) Comparison of protein precipitation methods for sample preparation prior to proteomic analysis. *Journal of Chromatography A* **1023**, 317-320

228. Singh, B., Sridhara, S., Arora, N., and Gangal, S. (1992) Evaluation of protein assay methods for pollen and fungal spore extracts. *Biochemistry international* **27**, 477-484

229. Fountoulakis, M., Juranville, J.-F., and Manneberg, M. (1992) Comparison of the Coomassie brilliant blue, bicinchoninic acid and Lowry quantitation assays, using non-glycosylated and glycosylated proteins. *Journal of biochemical and biophysical methods* **24**, 265-274

230. Helbig, A. O., Heck, A. J., and Slijper, M. (2010) Exploring the membrane proteome—challenges and analytical strategies. *Journal of proteomics* **73**, 868-878

231. Zhang, N., Chen, R., Young, N., Wishart, D., Winter, P., Weiner, J. H., and Li, L. (2007) Comparison of SDS-and methanol-assisted protein solubilization and digestion methods for Escherichia coli membrane proteome analysis by 2-D LC-MS/MS. *Proteomics* **7**, 484-493

232. Ishihama, Y., Rappsilber, J., and Mann, M. (2006) Modular Stop and Go Extraction Tips with Stacked Disks for Parallel and Multidimensional Peptide Fractionation in Proteomics. *Journal of Proteome Research* **5**, 988-994

233. Brosch, M., Swamy, S., Hubbard, T., and Choudhary, J. (2008) Comparison of Mascot and X!Tandem Performance for Low and High Accuracy Mass Spectrometry and the Development of an Adjusted Mascot Threshold. *Molecular & Cellular Proteomics* **7**, 962-970

234. Perkins, D. N., Pappin, D. J., Creasy, D. M., and Cottrell, J. S. (1999) Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **20**, 3551-3567

235. Duncan, D. T., Craig, R., and Link, A. J. (2005) Parallel Tandem: A Program for Parallel Processing of Tandem Mass Spectra Using PVM or MPI and X!Tandem. *Journal of Proteome Research* **4**, 1842-1847

236. Wilkins, M. R., Gasteiger, E., Gooley, A. A., Herbert, B. R., Molloy, M. P., Binz, P.-A., Ou, K., Sanchez, J.-C., Bairoch, A., and Williams, K. L. (1999)

High-throughput mass spectrometric discovery of protein post-translational modifications. *Journal of molecular biology* **289**, 645-657

237. Handley, J. (2002) Product Review: Software for MS Protein identfication. *Analytical chemistry* **74**, 159 A-162 A

238. Kapp, E. A., Schütz, F., Connolly, L. M., Chakel, J. A., Meza, J. E., Miller, C. A., Fenyo, D., Eng, J. K., Adkins, J. N., and Omenn, G. S. (2005) An evaluation, comparison, and accurate benchmarking of several publicly available MS/MS search algorithms: sensitivity and specificity analysis. *Proteomics* **5**, 3475-3490

239. Balgley, B. M., Laudeman, T., Yang, L., Song, T., and Lee, C. S. (2007) Comparative Evaluation of Tandem MS Search Algorithms Using a Target-Decoy Search Strategy. *Molecular & Cellular Proteomics* **6**, 1599-1608

240. Graf, L., Craik, C. S., Patthy, A., Roczniak, S., Fletterick, R. J., and Rutter, W. J. (1987) Selective alteration of substrate specificity by replacement of aspartic acid-189 with lysine in the binding pocket of trypsin. *Biochemistry* **26**, 2616-2623

241. Zee, B. M., and Garcia, B. A. (2012) Discovery of lysine post-translational modifications through mass spectrometric detection. *Essays in biochemistry* **52**, 147-163

242. Cox, J., and Mann, M. (2009) Computational principles of determining and improving mass precision and accuracy for proteome measurements in an Orbitrap. *Journal of the American Society for Mass Spectrometry* **20**, 1477-1485

243. Tomizioli, M., Lazar, C., Brugière, S., Burger, T., Salvi, D., Gatto, L., Moyet, L., Breckels, L. M., Hesse, A.-M., Lilley, K. S., Seigneurin-Berny, D., Finazzi, G., Rolland, N., and Ferro, M. (2014) Deciphering Thylakoid Sub-compartments using a Mass Spectrometry-based Approach. *Molecular & Cellular Proteomics : MCP* **13**, 2147-2167

244. Dvořáková, L., Cvrčková, F., and Fischer, L. (2007) Analysis of the hybrid proline-rich protein families from seven plant species suggests rapid diversification of their sequences and expression patterns. *BMC genomics* **8**, 1

245. Rao, M. V., Hale, B. A., and Ormrod, D. P. (1995) Amelioration of Ozone-Induced Oxidative Damage in Wheat Plants Grown under High Carbon Dioxide (Role of Antioxidant Enzymes). *Plant Physiology* **109**, 421-432

246. Zhu, M., Zhu, N., Song, W.-y., Harmon, A. C., Assmann, S. M., and Chen, S. (2014) Thiol-based Redox Proteins in Brassica napus Guard Cell Abscisic Acid and Methyl Jasmonate Signaling. *The Plant journal : for cell and molecular biology* **78**, 491-515

247. Yi, X., Hargett, S. R., Frankel, L. K., and Bricker, T. M. (2006) The PsbQ protein is required in Arabidopsis for photosystem II assembly/stability and photoautotrophy under low light conditions. *Journal Biological Chemistry* **281**, 26260-26267

248. Stemler, A. (1980) Forms of Dissolved Carbon Dioxide Required for Photosystem II Activity in Chloroplast Membranes. *Plant Physiol* **65**, 1160-1165

249. Okuda, K., Hammani, K., Tanz, S. K., Peng, L., Fukao, Y., Myouga, F., Motohashi, R., Shinozaki, K., Small, I., and Shikanai, T. (2010) The

pentatricopeptide repeat protein OTP82 is required for RNA editing of plastid ndhB and ndhG transcripts. *The Plant Journal* **61**, 339-349

250. Andrès, C., Agne, B., and Kessler, F. (2010) The TOC complex: Preprotein gateway to the chloroplast. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **1803**, 715-723

251. Agne, B., and Kessler, F. (2009) Protein transport in organelles: The Toc complex way of preprotein import. *The FEBS journal* **276**, 1156-1165

252. Choudhary, C., Kumar, C., Gnad, F., Nielsen, M. L., Rehman, M., Walther, T. C., Olsen, J. V., and Mann, M. (2009) Lysine acetylation targets protein complexes and co-regulates major cellular functions. *Science* **325**, 834-840

253. Mininno, M., Brugiere, S., Pautre, V., Gilgen, A., Ma, S., Ferro, M., Tardif, M., Alban, C., and Ravanel, S. (2012) Characterization of chloroplastic fructose 1,6-bisphosphate aldolases as lysine-methylated proteins in plants. *Journal Biological Chemistry* **287**, 21034-21044

254. Uematsu, K., Suzuki, N., Iwamae, T., Inui, M., and Yukawa, H. (2012) Increased fructose 1, 6-bisphosphate aldolase in plastids enhances growth and photosynthesis of tobacco plants. *Journal Experimental Botany* **63**, 3001-3009

255. Lu, W., Tang, X., Huo, Y., Xu, R., Qi, S., Huang, J., Zheng, C., and Wu, C.-a. (2012) Identification and characterization of fructose 1, 6-bisphosphate aldolase genes in Arabidopsis reveal a gene family with diverse responses to abiotic stresses. *Gene* **503**, 65-74

256. Rosano, G. L., and Ceccarelli, E. A. (2014) Recombinant protein expression in Escherichia coli: advances and challenges. *Recombinant protein expression in microbial systems*, 7

257. Hannig, G., and Makrides, S. C. (1998) Strategies for optimizing heterologous protein expression in Escherichia coli. *Trends in biotechnology* **16**, 54-60

258. Carnes, A. E., Hodgson, C. P., and Williams, J. A. (2006) Inducible Escherichia coli fermentation for increased plasmid DNA production. *Biotechnology and applied biochemistry* **45**, 155-166

259. Daber, R., Stayrook, S., Rosenberg, A., and Lewis, M. (2007) Structural Analysis of Lac Repressor Bound to Allosteric Effectors. *Journal of molecular biology* **370**, 609-619

260. Briñas, L., Zarazaga, M., Sáenz, Y., Ruiz-Larrea, F., and Torres, C. (2002) β-Lactamases in ampicillin-resistant Escherichia coli isolates from foods, humans, and healthy animals. *Antimicrobial agents and chemotherapy* **46**, 3156-3163

261. Kopito, R. R. (2000) Aggresomes, inclusion bodies and protein aggregation. *Trends in Cell Biology* **10**, 524-530

262. Sano, T., and Cantor, C. R. (1990) Expression of a cloned streptavidin gene in Escherichia coli. *PNAS* **87**, 142-146

263. Rudolph, R., and Lilie, H. (1996) In vitro folding of inclusion body proteins. *The FASEB Journal* **10**, 49-56

264. Oganesyan, N., Ankoudinova, I., Kim, S.-H., and Kim, R. (2007) Effect of osmotic stress and heat shock in recombinant protein overexpression and crystallization. *Protein expression and purification* **52**, 280-285

265. Capp, M. W., Pegram, L. M., Saecker, R. M., Kratz, M., Riccardi, D., Wendorff, T., Cannon, J. G., and Record Jr, M. T. (2009) Interactions of the osmolyte glycine betaine with molecular surfaces in water: thermodynamics, structural interpretation, and prediction of m-values. *Biochemistry* **48**, 10372-10379

266. Lucht, J. M., and Bremer, E. (1994) Adaptation of Escherichia coli to high osmolarity environments: osmoregulation of the high-affinity glycine betaine transport system ProU. *FEMS Microbiology reviews* **14**, 3-20

267. Liu, H., and Naismith, J. H. (2008) An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol. *BMC Biotechnology* **8**, 91-91

268. Portis, A. R., Salvucci, M. E., and Ogren, W. L. (1986) Activation of ribulosebisphosphate carboxylase/oxygenase at physiological CO2 and ribulosebisphosphate concentrations by Rubisco activase. *Plant Physiology* **82**, 967-971

269. Schürmann, M., and Sprenger, G. A. (2001) Fructose-6-phosphate aldolase is a novel class I aldolase from Escherichia coli and is related to a novel group of bacterial transaldolases. *Journal Biological Chemistry* **276**, 11055-11061

270. Hall, D. R., Leonard, G. A., Reed, C. D., Watt, C. I., Berry, A., and Hunter, W. N. (1999) The crystal structure of Escherichia coli class II fructose-1, 6-bisphosphate aldolase in complex with phosphoglycolohydroxamate reveals details of mechanism and specificity. *Journal of molecular biology* **287**, 383-394

271. Kelley, L. A., and Sternberg, M. J. (2009) Protein structure prediction on the Web: a case study using the Phyre server. *Nature protocols* **4**, 363-371

272. Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G., and McCoy, A. (2011) Overview of the CCP4 suite and current developments. *Acta Crystallographica Section D: Biological Crystallography* **67**, 235-242

273. Heo, L., Park, H., and Seok, C. (2013) GalaxyRefine: protein structure refinement driven by side-chain repacking. *Nucleic acids research* **41**, W384-W388

274. Schrödinger, L. The PyMOL Molecular Graphics System, Version 1.3 r1; Schrödinger, LLC: New York, 2010.

275. Costa, S., Almeida, A., Castro, A., and Domingues, L. (2014) Fusion tags for protein solubility, purification and immunogenicity in Escherichia coli: the novel Fh8 system. *Frontiers in Microbiology* **5**, 63

276. Seyferth, D., Dow, A. W., Menzel, H., and Flood, T. C. (1968) Trimethylsilyldiazomethane and trimethylsilylcarbene *Journal of the American Chemical Society* **90**, 1080-&

277. Rudolf, A. F., Skovgaard, T., Knapp, S., Jensen, L. J., and Berthelsen, J. (2014) A Comparison of Protein Kinases Inhibitor Screening Methods Using Both Enzymatic Activity and Binding Affinity Determination. *PLoS ONE* **9**, e98800

278. Eklund, T. (1984) The effect of carbon dioxide on bacterial growth and on uptake processes in bacterial membrane vesicles. *International Journal of Food Microbiology* **1**, 179-185

279. Kimura, R., and Miller, W. M. (1996) Effects of elevated pCO2 and/or osmolality on the growth and recombinant tPA production of CHO cells. *Biotechnology and bioengineering* **52**, 152-160

280. Curley, G. F., Laffey, J. G., and Kavanagh, B. P. (2013) Rebuttal from Gerard F. Curley, John G. Laffey and Brian P. Kavanagh. *The Journal of Physiology* **591**, 2771-2772

281. Beitler, J. R., Hubmayr, R. D., and Malhotra, A. (2013) CrossTalk opposing view: There is not added benefit to providing permissive hypercapnia in the treatment of ARDS. *The Journal of Physiology* **591**, 2767-2769

282. Domsic, J., Avvaru, B., Kim, C., Gruner, S., Agbandje-McKenna, M., Silverman, D., and McKenna, R. (2008) Entrapment of carbon dioxide in the active site of carbonic anhydrase II. *Journal Biological Chemistry* **283**, 30766-30771

283. Wiśniewski, J. R., Zougman, A., and Mann, M. (2009) Combination of FASP and StageTip-based fractionation allows in-depth analysis of the hippocampal membrane proteome. *Journal of proteome research* **8**, 5674-5678